THEORY OF FINANCIAL RISKS

FROM STATISTICAL PHYSICS TO RISK MANAGEMENT

JEAN-PHILIPPE BOUCHAUD and MARC POTTERS

Collection Aléa Saclay

CAMBRIDGE
UNIVERSITY PRESS

THEORY OF FINANCIAL RISKS

FROM STATISTICAL PHYSICS TO RISK MANAGEMENT

This book summarizes recent theoretical developments inspired by statistical physics in the description of the potential moves in financial markets, and its application to derivative pricing and risk control. The possibility of accessing and processing huge quantities of data on financial markets opens the path to new methodologies where systematic comparison between theories and real data not only becomes possible, but mandatory. This book takes a physicist's point of view of financial risk by comparing theory with experiment. Starting with important results in probability theory the authors discuss the statistical analysis of real data, the empirical determination of statistical laws, the definition of risk, the theory of optimal portfolio and the problem of derivatives (forward contracts, options). This book will be of interest to physicists interested in finance, quantitative analysts in financial institutions, risk managers and graduate students in mathematical finance.

JEAN-PHILIPPE BOUCHAUD was born in France in 1962. After studying at the French Lycée in London, he graduated from the École Normale Supérieure in Paris, where he also obtained his PhD in physics. He was then appointed by the CNRS until 1992, where he worked on diffusion in random media. After a year spent at the Cavendish Laboratory (Cambridge), Dr Bouchaud joined the Service de Physique de l'Etat Condensé (CEA-Saclay), where he works on the dynamics of glassy systems and on granular media. He became interested in theoretical finance in 1991 and founded the company Science & Finance in 1994 with J.-P. Aguilar. His work in finance includes extreme risk control and alternative option pricing models. He teaches statistical mechanics and finance in various *Grandes Écoles*. He was awarded the IBM young scientist prize in 1990 and the CNRS Silver Medal in 1996.

Born in Belgium in 1969, MARC POTTERS holds a PhD in physics from Princeton University and was a post-doctoral fellow at the University of Rome La Sapienza. In 1995, he joined Science & Finance, a research company located in Paris and founded by J.-P. Bouchaud and J.-P. Aguilar. Dr Potters is now Head of Research of S&F, supervising the work of six other physics PhDs. In collaboration with the researchers at S&F, he has published numerous articles in the new field of statistical finance and worked on concrete applications of financial forecasting, option pricing and risk control. Since 1998, he has also served as Head of Research of Capital Fund Management, a successful fund manager applying systematic trading strategies devised by S&F. Dr Potters teaches regularly with Dr Bouchaud at École Centrale de Paris.

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014, Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2000
Reprinted 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times 11/14pt. *System* LaTeX 2_ε [DBD]

A catalogue record of this book is available from the British Library

ISBN 0 521 78232 5 hardback

Contents

Foreword

Until recently, finance theory appeared to be reaching a triumphant climax. Many years ago, Harry Markowitz and William Sharpe had shown how diversification could reduce risk. In 1973, Fischer Black, Myron Scholes and Robert C. Merton went further by conjuring away risk completely, using the magic trick of dynamic replication. Twenty-five years later, a multi-trillion dollar derivatives industry had grown up around these insights. And of these five founding fathers, only Black missed out on a Nobel prize due to his tragic early death. Black, Scholes and Merton's option pricing breakthrough depended on the idea that hungry arbitrage traders were constantly prowling the markets, forcing prices to match theoretical predictions. The hedge fund Long-Term Capital Management – which included Scholes and Merton as partners – was founded with this principle at its core. So strong was LTCM's faith in these theories that it used leverage to make enormous bets on small discrepancies from the predictions of finance theory. We all know what happened next. In August and September 1998, the fund lost \$4.5 billion, roughly 90% of its value, and had to be bailed out by its 14 biggest counterparties. Global markets were severely disrupted for several months. All the shibboleths of finance theory, in particular diversification and replication, proved to be false gods, and the reputation of quants suffered badly as a result. Traditionally, finance texts take these shibboleths as a starting point, and build on them. Empirical verification is given scant attention, and the consequences of violating the key assumptions are often ignored completely. The result is a culture where markets get blamed if the theory breaks down, rather than vice versa, as it should be. Unsurprisingly, traders accuse some quants of having an ivory-tower mentality. Now, here come Bouchaud and Potters. Without eschewing rigour, they approach finance theory with a sceptical eye. All the familiar results – efficient portfolios, Black–Scholes and so on – are here, but with a strongly empirical flavour. There are also some useful additions to the existing toolkit, such as random matrix theory. Perhaps one day, theorists will show that the exact Black–Scholes regime is an unstable,

pathological state rather than the utopia it was formerly thought to be. Until then, quants will find this book a useful survival guide in the real world.

Nick Dunbar
Technical Editor, *Risk Magazine*
Author of *Inventing Money* (John Wiley and Sons, 2000)

Preface

Finance is a rapidly expanding field of science, with a rather unique link to applications. Correspondingly, recent years have witnessed the growing role of financial engineering in market rooms. The possibility of easily accessing and processing huge quantities of data on financial markets opens the path to new methodologies, where systematic comparison between theories and real data not only becomes possible, but mandatory. This perspective has spurred the interest of the statistical physics community, with the hope that methods and ideas developed in the past decades to deal with complex systems could also be relevant in finance. Correspondingly, many holders of PhDs in physics are now taking jobs in banks or other financial institutions.

However, the existing literature roughly falls into two categories: either rather abstract books from the mathematical finance community, which are very difficult for people trained in natural sciences to read, or more professional books, where the scientific level is usually quite poor.[1] In particular, there is in this context no book discussing the physicists' way of approaching scientific problems, in particular a systematic comparison between 'theory' and 'experiments' (i.e. empirical results), the art of approximations and the use of intuition.[2] Moreover, even in excellent books on the subject, such as the one by J. C. Hull, the point of view on derivatives is the traditional one of Black and Scholes, where the whole pricing methodology is based on the construction of *riskless strategies*. The idea of zero risk is counter-intuitive and the reason for the existence of these riskless strategies in the Black–Scholes theory is buried in the premises of Ito's stochastic differential rules.

It is our belief that a more intuitive understanding of these theories is needed for a better overall control of financial risks. The models discussed in *Theory of*

[1] There are notable exceptions, such as the remarkable book by J. C. Hull, *Futures, Options and Other Derivatives*, Prentice Hall, 1997.
[2] See however: I. Kondor, J. Kertesz (Eds): *Econophysics, an Emerging Science*, Kluwer, Dordrecht (1999); R. Mantegna and H. E. Stanley, *An Introduction to Econophysics*, Cambridge University Press (1999).

Financial Risk are devised to account for real markets' statistics where the construction of riskless hedges is in general impossible. The mathematical framework required to deal with these cases is however not more complicated, and has the advantage of making the issues at stake, in particular the problem of risk, more transparent.

Finally, commercial software packages are being developed to measure and control financial risks (some following the ideas developed in this book).[3] We hope that this book can be useful to all people concerned with financial risk control, by discussing at length the advantages and limitations of various statistical models.

Despite our efforts to remain simple, certain sections are still quite technical. We have used a smaller font to develop more advanced ideas, which are not crucial to understanding of the main ideas. Whole sections, marked by a star (*), contain rather specialized material and can be skipped at first reading. We have tried to be as precise as possible, but have sometimes been somewhat sloppy and non-rigorous. For example, the idea of probability is not axiomatized: its intuitive meaning is more than enough for the purpose of this book. The notation $P(\cdot)$ means the probability distribution for the variable which appears between the parentheses, and not a well-determined function of a dummy variable. The notation $x \to \infty$ does not necessarily mean that x tends to infinity in a mathematical sense, but rather that x is large. Instead of trying to derive results which hold true in any circumstances, we often compare order of magnitudes of the different effects: small effects are neglected, or included perturbatively.[4]

Finally, we have not tried to be comprehensive, and have left out a number of important aspects of theoretical finance. For example, the problem of interest rate derivatives (swaps, caps, swaptions...) is not addressed – we feel that the present models of interest rate dynamics are not satisfactory (see the discussion in Section 2.6). Correspondingly, we have not tried to give an exhaustive list of references, but rather to present our own way of understanding the subject. A certain number of important references are given at the end of each chapter, while more specialized papers are given as footnotes where we have found it necessary.

This book is divided into five chapters. Chapter 1 deals with important results in probability theory (the Central Limit Theorem and its limitations, the theory of extreme value statistics, etc.). The statistical analysis of real data, and the empirical determination of the statistical laws, are discussed in Chapter 2. Chapter 3 is concerned with the definition of risk, value-at-risk, and the theory of optimal

[3] For example, the software *Profiler*, commercialized by the company ATSM, heavily relies on the concepts introduced in Chapter 3.

[4] $a \simeq b$ means that a is of order b, $a \ll b$ means that a is smaller than, say, $b/10$. A computation neglecting terms of order $(a/b)^2$ is therefore accurate to 1%. Such a precision is usually enough in the financial context, where the uncertainty on the value of the parameters (such as the average return, the volatility, etc.), is often larger than 1%.

portfolio, in particular in the case where the probability of extreme risks has to be minimized. The problem of forward contracts and options, their optimal hedge and the residual risk is discussed in detail in Chapter 4. Finally, some more advanced topics on options are introduced in Chapter 5 (such as exotic options, or the role of transaction costs). Finally, a short glossary of financial terms, an index and a list of symbols are given at the end of the book, allowing one to find easily where each symbol or word was used and defined for the first time.

This book appeared in its first edition in French, under the title: *Théorie des Risques Financiers*, Aléa-Saclay-Eyrolles, Paris (1997). Compared to this first edition, the present version has been substantially improved and augmented. For example, we discuss the theory of random matrices and the problem of the interest rate curve, which were absent from the first edition. Furthermore, several points have been corrected or clarified.

Acknowledgements

This book owes a lot to discussions that we had with Rama Cont, Didier Sornette (who participated to the initial version of Chapter 3), and to the entire team of Science and Finance: Pierre Cizeau, Laurent Laloux, Andrew Matacz and Martin Meyer. We want to thank in particular Jean-Pierre Aguilar, who introduced us to the reality of financial markets, suggested many improvements, and supported us during the many years that this project took to complete. We also thank the companies ATSM and CFM, for providing financial data and for keeping us close to the real world. We also had many fruitful exchanges with Jeff Miller, and also with Alain Arnéodo, Aubry Miens,[5] Erik Aurell, Martin Baxter, Jean-François Chauwin, Nicole El Karoui, Stefano Galluccio, Gaëlle Gego, Giulia Iori, David Jeammet, Imre Kondor, Jean-Michel Lasry, Rosario Mantegna, Marc Mézard, Jean-François Muzy, Nicolas Sagna, Farhat Selmi, Gene Stanley, Ray Streater, Christian Walter, Mark Wexler and Karol Zyczkowski. We thank Claude Godrèche, who edited the French version of this book, for his friendly advice and support. Finally, J.-P.B. wants to thank Elisabeth Bouchaud for sharing so many far more important things.

This book is dedicated to our families, and more particularly to the memory of Paul Potters.

Paris, 1999

Jean-Philippe Bouchaud
Marc Potters

[5] With whom we discussed Eq. (1.24), which appears in his Diplomarbeit.

1

Probability theory: basic notions

All epistemologic value of the theory of probability is based on this: that large scale random phenomena in their collective action create strict, non random regularity.

(Gnedenko and Kolmogorov, *Limit Distributions for Sums of Independent Random Variables.*)

1.1 Introduction

Randomness stems from our incomplete knowledge of reality, from the lack of information which forbids a perfect prediction of the future. Randomness arises from complexity, from the fact that causes are diverse, that tiny perturbations may result in large effects. For over a century now, Science has abandoned Laplace's deterministic vision, and has fully accepted the task of deciphering randomness and inventing adequate tools for its description. The surprise is that, after all, randomness has many facets and that there are many levels to uncertainty, but, above all, that a new form of predictability appears, which is no longer deterministic but *statistical*.

Financial markets offer an ideal testing ground for these statistical ideas. The fact that a large number of participants, with divergent anticipations and conflicting interests, are simultaneously present in these markets, leads to an unpredictable behaviour. Moreover, financial markets are (sometimes strongly) affected by external news – which are, both in date and in nature, to a large degree unexpected. The statistical approach consists in drawing from past observations some information on the frequency of possible price changes. If one then assumes that these frequencies reflect some intimate mechanism of the markets themselves, then one may hope that these frequencies will remain stable in the course of time. For example, the mechanism underlying the roulette or the game of dice is obviously always the same, and one expects that the frequency of all possible

outcomes will be invariant in time – although of course each individual outcome is random.

This 'bet' that probabilities are stable (or better, stationary) is very reasonable in the case of roulette or dice;[1] it is nevertheless much less justified in the case of financial markets – despite the large number of participants which confer to the system a certain regularity, at least in the sense of Gnedenko and Kolmogorov. It is clear, for example, that financial markets do not behave now as they did 30 years ago: many factors contribute to the evolution of the way markets behave (development of derivative markets, world-wide and computer-aided trading, etc.). As will be mentioned in the following, 'young' markets (such as emergent countries markets) and more mature markets (exchange rate markets, interest rate markets, etc.) behave quite differently. The statistical approach to financial markets is based on the idea that whatever evolution takes place, this happens sufficiently *slowly* (on the scale of several years) so that the observation of the recent past is useful to describe a not too distant future. However, even this 'weak stability' hypothesis is sometimes badly in error, in particular in the case of a crisis, which marks a sudden change of market behaviour. The recent example of some Asian currencies indexed to the dollar (such as the Korean won or the Thai baht) is interesting, since the observation of past fluctuations is clearly of no help to predict the amplitude of the sudden turmoil of 1997, see Figure 1.1.

Hence, the statistical description of financial fluctuations is certainly imperfect. It is nevertheless extremely helpful: in practice, the 'weak stability' hypothesis is in most cases reasonable, at least to describe *risks*.[2]

In other words, the amplitude of the possible price changes (but not their sign!) is, to a certain extent, predictable. It is thus rather important to devise adequate tools, in order to *control* (if at all possible) financial risks. The goal of this first chapter is to present a certain number of basic notions in probability theory, which we shall find useful in the following. Our presentation does not aim at mathematical rigour, but rather tries to present the key concepts in an intuitive way, in order to ease their empirical use in practical applications.

1.2 Probabilities

1.2.1 Probability distributions

Contrarily to the throw of a dice, which can only return an integer between 1 and 6, the variation of price of a financial asset[3] can be arbitrary (we disregard

[1] The idea that science ultimately amounts to making the best possible guess of reality is due to R. P. Feynman (Seeking New Laws, in *The Character of Physical Laws*, MIT Press, Cambridge, MA, 1965).
[2] The prediction of *future returns* on the basis of past returns is however much less justified.
[3] *Asset* is the generic name for a financial instrument which can be bought or sold, like stocks, currencies, gold, bonds, etc.

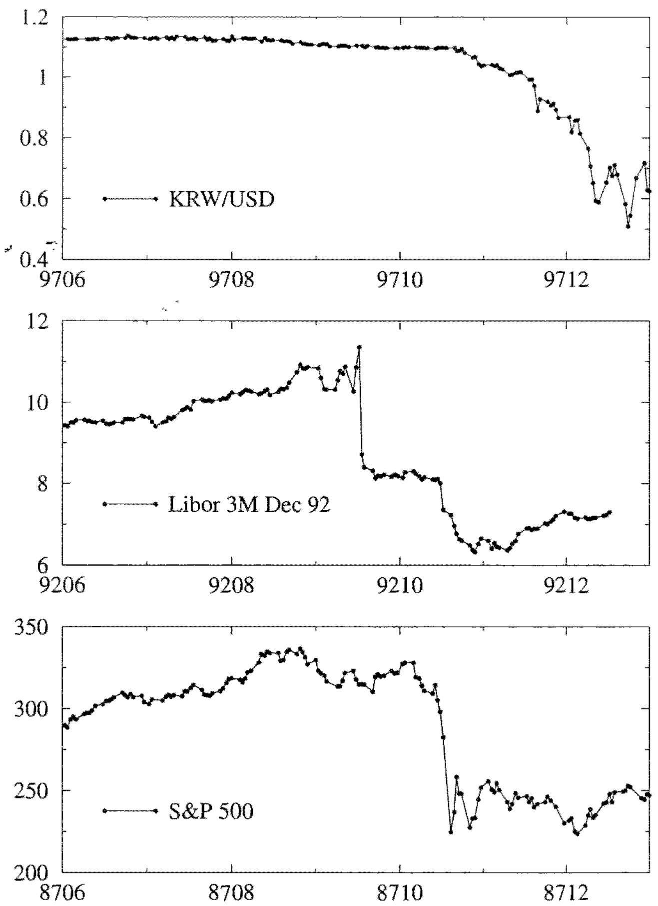

Fig. 1.1. Three examples of statistically unforeseen crashes: the Korean won against the dollar in 1997 (top), the British 3-month short-term interest rates futures in 1992 (middle), and the S&P 500 in 1987 (bottom). In the example of the Korean won, it is particularly clear that the distribution of price changes before the crisis was extremely narrow, and could not be extrapolated to anticipate what happened in the crisis period.

the fact that price changes cannot actually be smaller than a certain quantity – a 'tick'). In order to describe a random process X for which the result is a real number, one uses a probability density $P(x)$, such that the probability that X is within a small interval of width $\mathrm{d}x$ around $X = x$ is equal to $P(x)\,\mathrm{d}x$. In the following, we shall denote as $P(\cdot)$ the probability density for the variable appearing as the argument of the function. This is a potentially ambiguous, but very useful notation.

The probability that X is between a and b is given by the integral of $P(x)$ between a and b,

$$\mathcal{P}(a < X < b) = \int_a^b P(x)\,\mathrm{d}x. \tag{1.1}$$

In the following, the notation $\mathcal{P}(\cdot)$ means the probability of a given event, defined by the content of the parentheses $(\cdot)$.

The function $P(x)$ is a density; in this sense it depends on the units used to measure X. For example, if X is a length measured in centimetres, $P(x)$ is a probability density per unit length, i.e. per centimetre. The numerical value of $P(x)$ changes if X is measured in inches, but the probability that X lies between two specific values l_1 and l_2 is of course independent of the chosen unit. $P(x)\,\mathrm{d}x$ is thus invariant upon a change of unit, i.e. under the change of variable $x \to \gamma x$. More generally, $P(x)\,\mathrm{d}x$ is invariant upon any (monotonic) change of variable $x \to y(x)$: in this case, one has $P(x)\,\mathrm{d}x = P(y)\,\mathrm{d}y$.

In order to be a probability density in the usual sense, $P(x)$ must be non-negative ($P(x) \geq 0$ for all x) and must be normalized, that is that the integral of $P(x)$ over the whole range of possible values for X must be equal to one:

$$\int_{x_m}^{x_M} P(x)\,\mathrm{d}x = 1, \tag{1.2}$$

where x_m (resp. x_M) is the smallest value (resp. largest) which X can take. In the case where the possible values of X are not bounded from below, one takes $x_m = -\infty$, and similarly for x_M. One can actually always assume the bounds to be $\pm\infty$ by setting to zero $P(x)$ in the intervals $]-\infty, x_m]$ and $[x_M, \infty[$. Later in the text, we shall often use the symbol $\int$ as a shorthand for $\int_{-\infty}^{+\infty}$.

An equivalent way of describing the distribution of X is to consider its cumulative distribution $\mathcal{P}_<(x)$, defined as:

$$\mathcal{P}_<(x) \equiv \mathcal{P}(X < x) = \int_{-\infty}^{x} P(x')\,\mathrm{d}x'. \tag{1.3}$$

$\mathcal{P}_<(x)$ takes values between zero and one, and is monotonically increasing with x. Obviously, $\mathcal{P}_<(-\infty) = 0$ and $\mathcal{P}_<(+\infty) = 1$. Similarly, one defines $\mathcal{P}_>(x) = 1 - \mathcal{P}_<(x)$.

1.2.2 Typical values and deviations

It is quite natural to speak about 'typical' values of X. There are at least three mathematical definitions of this intuitive notion: the *most probable* value, the *median* and the *mean*. The most probable value x^* corresponds to the maximum of the function $P(x)$; x^* needs not be unique if $P(x)$ has several equivalent maxima.

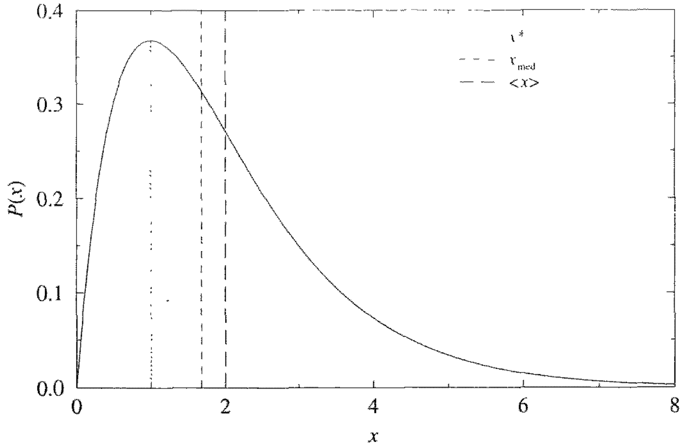

Fig. 1.2. The 'typical value' of a random variable X drawn according to a distribution density $P(x)$ can be defined in at least three different ways: through its mean value $\langle x \rangle$, its most probable value x^* or its median x_{med}. In the general case these three values are distinct.

The median x_{med} is such that the probabilities that X be greater or less than this particular value are equal. In other words, $\mathcal{P}_<(x_{\mathrm{med}}) = \mathcal{P}_>(x_{\mathrm{med}}) = \frac{1}{2}$. The mean, or *expected value* of X, which we shall note as m or $\langle x \rangle$ in the following, is the average of all possible values of X, weighted by their corresponding probability:

$$m \equiv \langle x \rangle = \int x P(x)\,\mathrm{d}x. \tag{1.4}$$

For a unimodal distribution (unique maximum), symmetrical around this maximum, these three definitions coincide. However, they are in general different, although often rather close to one another. Figure 1.2 shows an example of a non-symmetric distribution, and the relative position of the most probable value, the median and the mean.

One can then describe the fluctuations of the random variable X: if the random process is repeated several times, one expects the results to be scattered in a cloud of a certain 'width' in the region of typical values of X. This width can be described by the *mean absolute deviation* (MAD) E_{abs}, by the *root mean square* (RMS) σ (or, in financial terms, the *volatility*), or by the 'full width at half maximum' $w_{1/2}$.

The mean absolute deviation from a given reference value is the average of the distance between the possible values of X and this reference value,[4]

$$E_{\text{abs}} \equiv \int |x - x_{\text{med}}| P(x)\, \mathrm{d}x. \tag{1.5}$$

Similarly, the *variance* (σ^2) is the mean distance squared to the reference value m,

$$\sigma^2 \equiv \langle (x-m)^2 \rangle = \int (x-m)^2 P(x)\, \mathrm{d}x. \tag{1.6}$$

Since the variance has the dimension of x squared, its square root (the RMS, σ) gives the order of magnitude of the fluctuations around m.

Finally, the full width at half maximum $w_{1/2}$ is defined (for a distribution which is symmetrical around its unique maximum x^*) such that $P(x^* \pm (w_{1/2})/2) = P(x^*)/2$, which corresponds to the points where the probability density has dropped by a factor of two compared to its maximum value. One could actually define this width slightly differently, for example such that the total probability to find an event outside the interval $[(x^* - w/2), (x^* + w/2)]$ is equal to, say, 0.1.

The pair mean–variance is actually much more popular than the pair median–MAD. This comes from the fact that the absolute value is not an analytic function of its argument, and thus does not possess the nice properties of the variance, such as additivity under convolution, which we shall discuss below. However, for the empirical study of fluctuations, it is sometimes preferable to use the MAD; it is more *robust* than the variance, that is, less sensitive to rare extreme events, which may be the source of large statistical errors.

1.2.3 Moments and characteristic function

More generally, one can define higher-order *moments* of the distribution $P(x)$ as the average of powers of X:

$$m_n \equiv \langle x^n \rangle = \int x^n P(x)\, \mathrm{d}x. \tag{1.7}$$

Accordingly, the mean m is the first moment ($n = 1$), and the variance is related to the second moment ($\sigma^2 = m_2 - m^2$). The above definition, Eq. (1.7), is only meaningful if the integral converges, which requires that $P(x)$ decreases sufficiently rapidly for large $|x|$ (see below).

From a theoretical point of view, the moments are interesting: if they exist, their knowledge is often equivalent to the knowledge of the distribution $P(x)$ itself.[5] In

[4] One chooses as a reference value the median for the MAD and the mean for the RMS, because for a fixed distribution $P(x)$, these two quantities minimize, respectively, the MAD and the RMS.

[5] This is not rigorously correct, since one can exhibit examples of different distribution densities which possess exactly the same moments, see Section 1.3.2 below.

practice however, the high order moments are very hard to determine satisfactorily: as n grows, longer and longer time series are needed to keep a certain level of precision on m_n; these high moments are thus in general not adapted to describe empirical data.

For many computational purposes, it is convenient to introduce the *characteristic function* of $P(x)$, defined as its Fourier transform:

$$\hat{P}(z) \equiv \int \mathrm{e}^{\mathrm{i}zx} P(x)\, \mathrm{d}x. \tag{1.8}$$

The function $P(x)$ is itself related to its characteristic function through an inverse Fourier transform:

$$P(x) = \frac{1}{2\pi} \int \mathrm{e}^{-\mathrm{i}zx} \hat{P}(z)\, \mathrm{d}z. \tag{1.9}$$

Since $P(x)$ is normalized, one always has $\hat{P}(0) = 1$. The moments of $P(x)$ can be obtained through successive derivatives of the characteristic function at $z = 0$,

$$m_n = (-\mathrm{i})^n \left. \frac{\mathrm{d}^n}{\mathrm{d}z^n} \hat{P}(z) \right|_{z=0}. \tag{1.10}$$

One finally defines the *cumulants* c_n of a distribution as the successive derivatives of the logarithm of its characteristic function:

$$c_n = (-\mathrm{i})^n \left. \frac{\mathrm{d}^n}{\mathrm{d}z^n} \log \hat{P}(z) \right|_{z=0}. \tag{1.11}$$

The cumulant c_n is a polynomial combination of the moments m_p with $p \leq n$. For example $c_2 = m_2 - m^2 = \sigma^2$. It is often useful to normalize the cumulants by an appropriate power of the variance, such that the resulting quantities are dimensionless. One thus defines the *normalized cumulants* λ_n,

$$\lambda_n \equiv c_n / \sigma^n. \tag{1.12}$$

One often uses the third and fourth normalized cumulants, called the *skewness* and *kurtosis* (κ),[6]

$$\lambda_3 = \frac{\langle (x-m)^3 \rangle}{\sigma^3} \qquad \kappa \equiv \lambda_4 = \frac{\langle (x-m)^4 \rangle}{\sigma^4} - 3. \tag{1.13}$$

The above definition of cumulants may look arbitrary, but these quantities have remarkable properties. For example, as we shall show in Section 1.5, the cumulants simply add when one sums independent random variables. Moreover a Gaussian distribution (or the normal law of Laplace and Gauss) is characterized by the fact that all cumulants of order larger than two are identically zero. Hence the

[6] Note that it is sometimes $\kappa + 3$, rather than κ itself, which is called the kurtosis.

cumulants, in particular κ, can be interpreted as a measure of the distance between a given distribution $P(x)$ and a Gaussian.

1.2.4 Divergence of moments – asymptotic behaviour

The moments (or cumulants) of a given distribution do not always exist. A necessary condition for the nth moment (m_n) to exist is that the distribution density $P(x)$ should decay faster than $1/|x|^{n+1}$ for $|x|$ going towards infinity, or else the integral, Eq. (1.7), would diverge for $|x|$ large. If one only considers distribution densities that are behaving asymptotically as a power-law, with an exponent $1+\mu$,

$$P(x) \sim \frac{\mu A_{\pm}^{\mu}}{|x|^{1+\mu}} \text{ for } x \to \pm\infty, \tag{1.14}$$

then all the moments such that $n \geq \mu$ are infinite. For example, such a distribution has no finite variance whenever $\mu \leq 2$. [Note that, for $P(x)$ to be a normalizable probability distribution, the integral, Eq. (1.2), must converge, which requires $\mu > 0$.]

The characteristic function of a distribution having an asymptotic power-law behaviour given by Eq. (1.14) is non-analytic around $z = 0$. The small z expansion contains regular terms of the form z^n for $n < \mu$ followed by a non-analytic term $|z|^{\mu}$ (possibly with logarithmic corrections such as $|z|^{\mu} \log z$ for integer μ). The derivatives of order larger or equal to μ of the characteristic function thus do not exist at the origin ($z = 0$).

1.3 Some useful distributions

1.3.1 Gaussian distribution

The most commonly encountered distributions are the 'normal' laws of Laplace and Gauss, which we shall simply call Gaussian in the following. Gaussians are ubiquitous: for example, the number of *heads* in a sequence of a thousand coin tosses, the exact number of oxygen molecules in the room, the height (in inches) of a randomly selected individual, are all approximately described by a Gaussian distribution.[7] The ubiquity of the Gaussian can be in part traced to the Central Limit Theorem (CLT) discussed at length below, which states that a phenomenon resulting from a large number of small independent causes is Gaussian. There exists however a large number of cases where the distribution describing a complex phenomenon is *not* Gaussian: for example, the amplitude of earthquakes, the velocity differences in a turbulent fluid, the stresses in granular materials, etc., and, as we shall discuss in the next chapter, the price fluctuations of most financial assets.

[7] Although, in the above three examples, the random variable cannot be negative. As we shall discuss below, the Gaussian description is generally only valid in a certain neighbourhood of the maximum of the distribution.

A Gaussian of mean m and root mean square σ is defined as:

$$P_G(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right). \tag{1.15}$$

The median and most probable value are in this case equal to m, whereas the MAD (or any other definition of the width) is proportional to the RMS (for example, $E_{\text{abs}} = \sigma\sqrt{2/\pi}$). For $m = 0$, all the odd moments are zero and the even moments are given by $m_{2n} = (2n-1)(2n-3)\ldots\sigma^{2n} = (2n-1)!!\,\sigma^{2n}$.

All the cumulants of order greater than two are zero for a Gaussian. This can be realized by examining its characteristic function:

$$\hat{P}_G(z) = \exp\left(-\frac{\sigma^2 z^2}{2} + imz\right). \tag{1.16}$$

Its logarithm is a second-order polynomial, for which all derivatives of order larger than two are zero. In particular, the kurtosis of a Gaussian variable is zero. As mentioned above, the kurtosis is often taken as a measure of the distance from a Gaussian distribution. When $\kappa > 0$ (*leptokurtic* distributions), the corresponding distribution density has a marked peak around the mean, and rather 'thick' tails. Conversely, when $\kappa < 0$, the distribution density has a flat top and very thin tails. For example, the uniform distribution over a certain interval (for which tails are absent) has a kurtosis $\kappa = -\frac{6}{5}$.

A Gaussian variable is peculiar because 'large deviations' are extremely rare. The quantity $\exp(-x^2/2\sigma^2)$ decays so fast for large x that deviations of a few times σ are nearly impossible. For example, a Gaussian variable departs from its most probable value by more than 2σ only 5% of the times, of more than 3σ in 0.2% of the times, whereas a fluctuation of 10σ has a probability of less than 2×10^{-23}; in other words, it never happens.

1.3.2 Log-normal distribution

Another very popular distribution in mathematical finance is the so-called 'log-normal' law. That X is a log-normal random variable simply means that $\log X$ is normal, or Gaussian. Its use in finance comes from the assumption that the *rate of returns*, rather than the absolute change of prices, are independent random variables. The increments of the logarithm of the price thus asymptotically sum to a Gaussian, according to the CLT detailed below. The log-normal distribution

density is thus defined as:[8]

$$P_{\mathrm{LN}}(x) \equiv \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\log^2(x/x_0)}{2\sigma^2}\right), \tag{1.17}$$

the moments of which being: $m_n = x_0^n e^{n^2\sigma^2/2}$.

In the context of mathematical finance, one often prefers log-normal to Gaussian distributions for several reasons. As mentioned above, the existence of a random rate of return, or random interest rate, naturally leads to log-normal statistics. Furthermore, log-normals account for the following symmetry in the problem of exchange rates:[9] if x is the rate of currency A in terms of currency B, then obviously, $1/x$ is the rate of currency B in terms of A. Under this transformation, $\log x$ becomes $-\log x$ and the description in terms of a log-normal distribution (or in terms of any other even function of $\log x$) is independent of the reference currency. One often hears the following argument in favour of log-normals: since the price of an asset cannot be negative, its statistics cannot be Gaussian since the latter admits in principle negative values, whereas a log-normal excludes them by construction. This is however a red-herring argument, since the description of the fluctuations of the price of a financial asset in terms of Gaussian or log-normal statistics is in any case an *approximation* which is only valid in a certain range. As we shall discuss at length below, these approximations are totally unadapted to describe extreme risks. Furthermore, even if a price drop of more than 100% is in principle possible for a Gaussian process,[10] the error caused by neglecting such an event is much smaller than that induced by the use of either of these two distributions (Gaussian or log-normal). In order to illustrate this point more clearly, consider the probability of observing n times 'heads' in a series of N coin tosses, which is exactly equal to $2^{-N}C_N^n$. It is also well known that in the neighbourhood of $N/2$, $2^{-N}C_N^n$ is very accurately approximated by a Gaussian of variance $N/4$; this is however not contradictory with the fact that $n \geq 0$ by construction!

Finally, let us note that for moderate volatilities (up to say 20%), the two distributions (Gaussian and log-normal) look rather alike, especially in the 'body' of the distribution (Fig. 1.3). As for the tails, we shall see below that Gaussians substantially underestimate their weight, whereas the log-normal predicts that large

8 A log-normal distribution has the remarkable property that the knowledge of all its moments is not sufficient to characterize the corresponding distribution. It is indeed easy to show that the following distribution: $\frac{1}{\sqrt{2\pi}}x^{-1}\exp\left[-\frac{1}{2}(\log x)^2\right][1 + a\sin(2\pi\log x)]$, for $|a| \leq 1$, has moments which are independent of the value of a, and thus coincide with those of a log-normal distribution, which corresponds to $a = 0$ ([Feller], p. 227).

9 This symmetry is however not always obvious. The dollar, for example, plays a special role. This symmetry can only be expected between currencies of similar strength.

10 In the rather extreme case of a 20% annual volatility and a zero annual return, the probability for the price to become negative after a year in a Gaussian description is less than one out of 3 million.

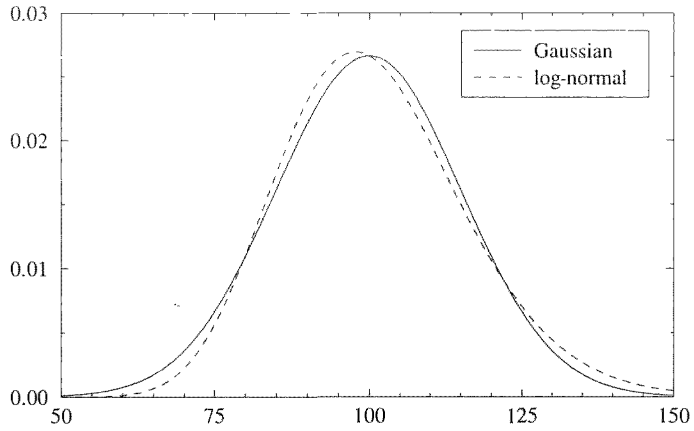

Fig. 1.3. Comparison between a Gaussian (thick line) and a log-normal (dashed line), with $m = x_0 = 100$ and σ equal to 15 and 15% respectively. The difference between the two curves shows up in the tails.

positive jumps are more frequent than large negative jumps. This is at variance with empirical observation: the distributions of absolute stock price changes are rather symmetrical; if anything, large negative draw-downs are more frequent than large positive draw-ups.

1.3.3 *Lévy distributions and Paretian tails*

Lévy distributions (noted $L_\mu(x)$ below) appear naturally in the context of the CLT (see below), because of their stability property under addition (a property shared by Gaussians). The tails of Lévy distributions are however much 'fatter' than those of Gaussians, and are thus useful to describe multiscale phenomena (i.e. when both very large and very small values of a quantity can commonly be observed – such as personal income, size of pension funds, amplitude of earthquakes or other natural catastrophes, etc.). These distributions were introduced in the 1950s and 1960s by Mandelbrot (following Pareto) to describe personal income and the price changes of some financial assets, in particular the price of cotton [Mandelbrot]. An important constitutive property of these Lévy distributions is their power-law behaviour for large arguments, often called 'Pareto tails':

$$L_\mu(x) \sim \frac{\mu A_\pm^\mu}{|x|^{1+\mu}} \text{ for } x \to \pm\infty, \tag{1.18}$$

where $0 < \mu < 2$ is a certain exponent (often called α), and $A_\pm^\mu$ two constants which we call *tail amplitudes*, or *scale parameters*: $A_\pm$ indeed gives the order of

magnitude of the large (positive or negative) fluctuations of x. For instance, the probability to draw a number larger than x decreases as $\mathcal{P}_>(x) = (A_+/x)^\mu$ for large positive x.

One can of course in principle observe Pareto tails with $\mu \geq 2$; but, those tails do not correspond to the asymptotic behaviour of a Lévy distribution.

In full generality, Lévy distributions are characterized by an *asymmetry parameter* defined as $\beta \equiv (A_+^\mu - A_-^\mu)/(A_+^\mu + A_-^\mu)$, which measures the relative weight of the positive and negative tails. We shall mostly focus in the following on the symmetric case $\beta = 0$. The fully asymmetric case ($\beta = 1$) is also useful to describe strictly positive random variables, such as, for example, the time during which the price of an asset remains below a certain value, etc.

An important consequence of Eq. (1.14) with $\mu \leq 2$ is that the variance of a Lévy distribution is formally infinite: the probability density does not decay fast enough for the integral, Eq. (1.6), to converge. In the case $\mu \leq 1$, the distribution density decays so slowly that even the mean, or the MAD, fail to exist.[11] The scale of the fluctuations, defined by the width of the distribution, is always set by $A = A_+ = A_-$.

There is unfortunately no simple analytical expression for symmetric Lévy distributions $L_\mu(x)$, except for $\mu = 1$, which corresponds to a Cauchy distribution (or 'Lorentzian'):

$$L_1(x) = \frac{A}{x^2 + \pi^2 A^2}. \tag{1.19}$$

However, the characteristic function of a symmetric Lévy distribution is rather simple, and reads:

$$\hat{L}_\mu(z) = \exp\left(-a_\mu |z|^\mu\right), \tag{1.20}$$

where a_μ is a certain constant, proportional to the tail parameter A^μ.[12] It is thus clear that in the limit $\mu = 2$, one recovers the definition of a Gaussian. When μ decreases from 2, the distribution becomes more and more sharply peaked around the origin and fatter in its tails, while 'intermediate' events lose weight (Fig. 1.4). These distributions thus describe 'intermittent' phenomena, very often small, sometimes gigantic.

Note finally that Eq. (1.20) does not define a probability distribution when $\mu > 2$, because its inverse Fourier transform is not everywhere positive.

In the case $\beta \neq 0$, one would have:

$$\hat{L}_\mu^\beta(z) = \exp\left[-a_\mu |z|^\mu \left(1 + i\beta \tan(\mu\pi/2)\frac{z}{|z|}\right)\right] \quad (\mu \neq 1). \tag{1.21}$$

[11] The median and the most probable value however still exist. For a symmetric Lévy distribution, the most probable value defines the so-called 'localization' parameter m.

[12] For example, when $1 < \mu < 2$, $A^\mu = \mu\Gamma(\mu - 1)\sin(\pi\mu/2)a_\mu/\pi$.

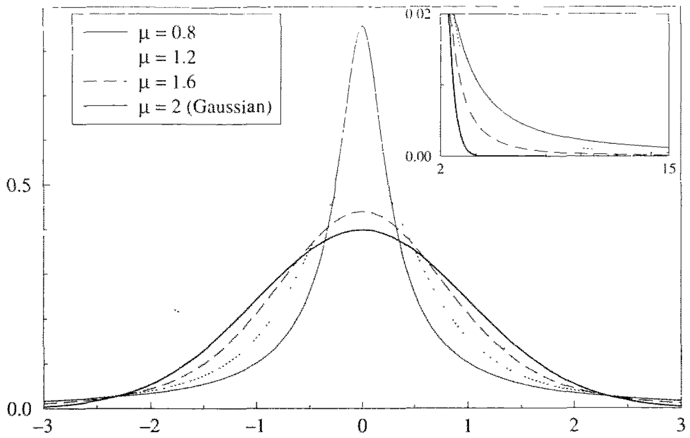

Fig. 1.4. Shape of the symmetric Lévy distributions with $\mu = 0.8$, 1.2, 1.6 and 2 (this last value actually corresponds to a Gaussian). The smaller μ, the sharper the 'body' of the distribution, and the fatter the tails, as illustrated in the inset.

It is important to notice that while the leading asymptotic term for large x is given by Eq. (1.18), there are subleading terms which can be important for finite x. The full asymptotic series actually reads:

$$L_\mu(x) = \sum_{n=1}^{\infty} \frac{(-)^{n+1}}{\pi n!} \frac{a_\mu^n}{x^{1+n\mu}} \Gamma(1 + n\mu) \sin(\pi\mu n/2). \tag{1.22}$$

The presence of the subleading terms may lead to a bad empirical estimate of the exponent μ based on a fit of the tail of the distribution. In particular, the 'apparent' exponent which describes the function L_μ for finite x is larger than μ, and decreases towards μ for $x \to \infty$, but more and more slowly as μ gets nearer to the Gaussian value $\mu = 2$, for which the power-law tails no longer exist. Note however that one also often observes empirically the opposite behaviour, i.e. an apparent Pareto exponent which *grows* with x. This arises when the Pareto distribution, Eq. (1.18), is only valid in an intermediate regime $x \ll 1/\alpha$, beyond which the distribution decays exponentially, say as $\exp(-\alpha x)$. The Pareto tail is then 'truncated' for large values of x, and this leads to an effective μ which grows with x.

An interesting generalization of the Lévy distributions which accounts for this exponential cut-off is given by the 'truncated Lévy distributions' (TLD), which will be of much use in the following. A simple way to alter the characteristic function

Eq. (1.20) to account for an exponential cut-off for large arguments is to set:[13]

$$\hat{L}_{\mu}^{(t)}(z) = \exp\left[-a_{\mu}\frac{(\alpha^2+z^2)^{\frac{\mu}{2}}\cos\left(\mu\arctan(|z|/\alpha)\right)-\alpha^{\mu}}{\cos(\pi\mu/2)}\right], \tag{1.23}$$

for $1 \le \mu \le 2$. The above form reduces to Eq. (1.20) for $\alpha = 0$. Note that the argument in the exponential can also be written as:

$$\frac{a_{\mu}}{2\cos(\pi\mu/2)}\left[(\alpha+\mathrm{i}z)^{\mu}+(\alpha-\mathrm{i}z)^{\mu}-2\alpha^{\mu}\right]. \tag{1.24}$$

Exponential tail: a limiting case

Very often in the following, we shall notice that in the formal limit $\mu \to \infty$, the power-law tail becomes an exponential tail, if the tail parameter is simultaneously scaled as $A^{\mu} = (\mu/\alpha)^{\mu}$. Qualitatively, this can be understood as follows: consider a probability distribution restricted to positive x, which decays as a power-law for large x, defined as:

$$\mathcal{P}_{>}(x) = \frac{A^{\mu}}{(A+x)^{\mu}}. \tag{1.25}$$

This shape is obviously compatible with Eq. (1.18), and is such that $\mathcal{P}_{>}(x=0) = 1$. If $A = (\mu/\alpha)$, one then finds:

$$\mathcal{P}_{>}(x) = \frac{1}{[1+(\alpha x/\mu)]^{\mu}} \underset{\mu\to\infty}{\longrightarrow} \exp(-\alpha x). \tag{1.26}$$

1.3.4 Other distributions (*)

There are obviously a very large number of other statistical distributions useful to describe random phenomena. Let us cite a few, which often appear in a financial context:

- The discrete Poisson distribution: consider a set of points randomly scattered on the real axis, with a certain density ω (e.g. the times when the price of an asset changes). The number of points n in an arbitrary interval of length ℓ is distributed according to the Poisson distribution:

$$P(n) \equiv \frac{(\omega\ell)^n}{n!}\exp(-\omega\ell). \tag{1.27}$$

- The hyperbolic distribution, which interpolates between a Gaussian 'body' and exponential tails:

$$P_{\mathrm{H}}(x) \equiv \frac{1}{2x_0 K_1(\alpha x_0)}\exp-[\alpha\sqrt{x_0^2+x^2}], \tag{1.28}$$

where the normalization $K_1(\alpha x_0)$ is a modified Bessel function of the second

[13] See I. Koponen, Analytic approach to the problem of convergence to truncated Lévy flights towards the Gaussian stochastic process, *Physical Review E*, **52**, 1197 (1995).

kind. For x small compared to x_0, $P_{\mathrm{H}}(x)$ behaves as a Gaussian although its asymptotic behaviour for $x \gg x_0$ is fatter and reads $\exp(-\alpha|x|)$.

From the characteristic function

$$\hat{P}_{\mathrm{H}}(z) = \frac{\alpha x_0 K_1(x_0\sqrt{1+\alpha z})}{K_1(\alpha x_0)\sqrt{1+\alpha z}}, \tag{1.29}$$

we can compute the variance

$$\sigma^2 = \frac{x_0 K_2(\alpha x_0)}{\alpha K_1(\alpha x_0)}, \tag{1.30}$$

and kurtosis

$$\kappa = 3\left(\frac{K_2(\alpha x_0)}{K_1(\alpha x_0)}\right)^2 + \frac{12}{\alpha x_0}\frac{K_2(\alpha x_0)}{K_1(\alpha x_0)} - 3. \tag{1.31}$$

Note that the kurtosis of the hyperbolic distribution is always between zero and three. In the case $x_0 = 0$, one finds the symmetric exponential distribution:

$$P_{\mathrm{E}}(x) = \frac{\alpha}{2}\exp(-\alpha|x|), \tag{1.32}$$

with even moments $m_{2n} = (2n)!\,\alpha^{-2n}$, which gives $\sigma^2 = 2\alpha^{-2}$ and $\kappa = 3$. Its characteristic function reads: $\hat{P}_{\mathrm{E}}(z) = \alpha^2/(\alpha^2+z^2)$.

- The Student distribution, which also has power-law tails:

$$P_{\mathrm{S}}(x) \equiv \frac{1}{\sqrt{\pi}}\frac{\Gamma((1+\mu)/2)}{\Gamma(\mu/2)}\frac{a^{\mu}}{(a^2+x^2)^{(1+\mu)/2}}, \tag{1.33}$$

which coincides with the Cauchy distribution for $\mu = 1$, and tends towards a Gaussian in the limit $\mu \to \infty$, provided that a^2 is scaled as μ. The even moments of the Student distribution read: $m_{2n} = (2n-1)!!\,\Gamma(\mu/2-n)/\Gamma(\mu/2)\,(a^2/2)^n$, provided $2n < \mu$; and are infinite otherwise. One can check that in the limit $\mu \to \infty$, the above expression gives back the moments of a Gaussian: $m_{2n} = (2n-1)!!\,\sigma^{2n}$. Figure 1.5 shows a plot of the Student distribution with $\kappa = 1$, corresponding to $\mu = 10$.

1.4 Maximum of random variables – statistics of extremes

If one observes a series of N independent realizations of the same random phenomenon, a question which naturally arises, in particular when one is concerned about risk control, is to determine the order of magnitude of the *maximum* observed value of the random variable (which can be the price drop of a financial asset, or the water level of a flooding river, etc.). For example, in Chapter 3, the so-called 'value-at-risk' (VaR) on a typical horizon will be defined as the possible maximum loss over that period (within a certain confidence level).

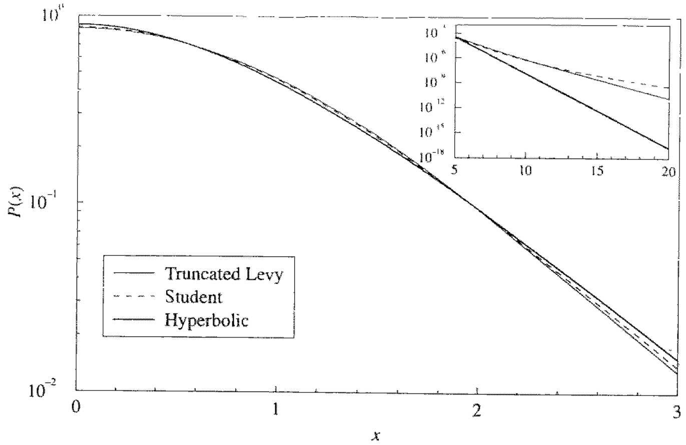

Fig. 1.5. Probability density for the truncated Lévy ($\mu = \frac{3}{2}$), Student and hyperbolic distributions. All three have two free parameters which were fixed to have unit variance and kurtosis. The inset shows a blow-up of the tails where one can see that the Student distribution has tails similar to (but slightly thicker than) those of the truncated Lévy.

The law of large numbers tells us that an event which has a probability p of occurrence appears on average Np times on a series of N observations. One thus expects to observe events which have a probability of at least $1/N$. It would be surprising to encounter an event which has a probability much smaller than $1/N$. The order of magnitude of the largest event, $\Lambda_{\max}$, observed in a series of N independent identically distributed (iid) random variables is thus given by:

$$\mathcal{P}_>(\Lambda_{\max}) = 1/N. \tag{1.34}$$

More precisely, the full probability distribution of the maximum value $x_{\max} = \max_{i=1,N}\{x_i\}$, is relatively easy to characterize; this will justify the above simple criterion Eq. (1.34). The cumulative distribution $\mathcal{P}(x_{\max} < \Lambda)$ is obtained by noticing that if the maximum of all x_i's is smaller than Λ, all of the x_i's must be smaller than Λ. If the random variables are iid, one finds:

$$\mathcal{P}(x_{\max} < \Lambda) = [\mathcal{P}_<(\Lambda)]^N. \tag{1.35}$$

Note that this result is general, and does not rely on a specific choice for $P(x)$. When Λ is large, it is useful to use the following approximation:

$$\mathcal{P}(x_{\max} < \Lambda) = [1 - \mathcal{P}_>(\Lambda)]^N \simeq e^{-N\mathcal{P}_>(\Lambda)}. \tag{1.36}$$

Since we now have a simple formula for the distribution of $x_{\max}$, one can invert

it in order to obtain, for example, the median value of the maximum, noted Λ_{med}, such that $\mathcal{P}(x_{\max} < \Lambda_{\text{med}}) = \frac{1}{2}$:

$$\mathcal{P}_>(\Lambda_{\text{med}}) = 1 - \left(\tfrac{1}{2}\right)^{1/N} \simeq \frac{\log 2}{N}. \tag{1.37}$$

More generally, the value Λ_p which is greater than $x_{\max}$ with probability p is given by

$$\mathcal{P}_>(\Lambda_p) \simeq -\frac{\log p}{N}. \tag{1.38}$$

The quantity $\Lambda_{\max}$ defined by Eq. (1.34) above is thus such that $p = 1/e \simeq 0.37$. The probability that $x_{\max}$ is even *larger* than $\Lambda_{\max}$ is thus 63%. As we shall now show, $\Lambda_{\max}$ also corresponds, in many cases, to the *most probable value* of $x_{\max}$.

Equation (1.38) will be very useful in Chapter 3 to estimate a maximal potential loss within a certain confidence level. For example, the largest daily loss Λ expected next year, with 95% confidence, is defined such that $\mathcal{P}_<(-\Lambda) = -\log(0.95)/250$, where $\mathcal{P}_<$ is the cumulative distribution of daily price changes, and 250 is the number of market days per year.

Interestingly, the distribution of $x_{\max}$ only depends, when N is large, on the asymptotic behaviour of the distribution of x, $P(x)$, when $x \to \infty$. For example, if $P(x)$ behaves as an exponential when $x \to \infty$, or more precisely if $\mathcal{P}_>(x) \sim \exp(-\alpha x)$, one finds:

$$\Lambda_{\max} = \frac{\log N}{\alpha}, \tag{1.39}$$

which grows very slowly with N.[14] Setting $x_{\max} = \Lambda_{\max} + (u/\alpha)$, one finds that the deviation u around $\Lambda_{\max}$ is distributed according to the Gumbel distribution:

$$P(u) = e^{-e^{-u}} e^{-u}. \tag{1.40}$$

The most probable value of this distribution is $u = 0$.[15] This shows that $\Lambda_{\max}$ is the most probable value of $x_{\max}$. The result, Eq. (1.40), is actually much more general, and is valid as soon as $P(x)$ decreases more rapidly than any power-law for $x \to \infty$: the deviation between $\Lambda_{\max}$ (defined as Eq. (1.34)) and $x_{\max}$ is always distributed according to the Gumbel law, Eq. (1.40), up to a scaling factor in the definition of u.

The situation is radically different if $P(x)$ decreases as a power-law, cf. Eq. (1.14). In this case,

$$\mathcal{P}_>(x) \simeq \frac{A_+^\mu}{x^\mu}, \tag{1.41}$$

[14] For example, for a symmetric exponential distribution $P(x) = \exp(-|x|)/2$, the median value of the maximum of $N = 10\,000$ variables is only 6.3.

[15] This distribution is discussed further in the context of financial risk control in Section 3.1.2, and drawn in Figure 3.1.

and the typical value of the maximum is given by:

$$\Lambda_{\max} = A_+ N^{\frac{1}{\mu}}. \qquad (1.42)$$

Numerically, for a distribution with $\mu = \frac{3}{2}$ and a scale factor $A_+ = 1$, the largest of $N = 10\,000$ variables is on the order of 450, whereas for $\mu = \frac{1}{2}$ it is one hundred million! The complete distribution of the maximum, called the Fréchet distribution, is given by:

$$P(u) = \frac{\mu}{u^{1+\mu}} \mathrm{e}^{-1/u^{\mu}} \qquad u = \frac{x_{\max}}{A_+ N^{\frac{1}{\mu}}}. \qquad (1.43)$$

Its asymptotic behaviour for $u \to \infty$ is still a power-law of exponent $1 + \mu$. Said differently, both power-law tails and exponential tails *are stable with respect to the 'max' operation.*[16] The most probable value $x_{\max}$ is now equal to $(\mu/1+\mu)^{1/\mu}\Lambda_{\max}$. As mentioned above, the limit $\mu \to \infty$ formally corresponds to an exponential distribution. In this limit, one indeed recovers $\Lambda_{\max}$ as the most probable value.

Equation (1.42) allows us to discuss intuitively the divergence of the mean value for $\mu \le 1$ and of the variance for $\mu \le 2$. If the mean value exists, the sum of N random variables is typically equal to Nm, where m is the mean (see also below). But when $\mu < 1$, the largest encountered value of X is on the order of $N^{1/\mu} \gg N$, and would thus be larger than the entire sum. Similarly, as discussed below, when the variance exists, the RMS of the sum is equal to $\sigma\sqrt{N}$. But for $\mu < 2$, $x_{\max}$ grows faster than $\sqrt{N}$.

More generally, one can rank the random variables x_i in decreasing order, and ask for an estimate of the nth encountered value, noted $\Lambda[n]$ below. (In particular, $\Lambda[1] = x_{\max}$). The distribution P_n of $\Lambda[n]$ can be obtained in full generality as:

$$P_n(\Lambda[n]) = N C_{N-1}^{n-1}\, P(x = \Lambda[n])\, (\mathcal{P}(x > \Lambda[n])^{n-1} (\mathcal{P}(x < \Lambda[n])^{N-n}. \qquad (1.44)$$

The previous expression means that one has first to choose $\Lambda[n]$ among N variables (N ways), $n-1$ variables among the $N-1$ remaining as the $n-1$ largest ones (C_{N-1}^{n-1} ways), and then assign the corresponding probabilities to the configuration where $n-1$ of them are larger than $\Lambda[n]$ and $N-n$ are smaller than $\Lambda[n]$. One can study the position $\Lambda^*[n]$ of the maximum of P_n, and also the width of P_n, defined from the second derivative of $\log P_n$ calculated at $\Lambda^*[n]$. The calculation simplifies in the limit where $N \to \infty$, $n \to \infty$, with the ratio n/N fixed. In this limit, one finds a relation which generalizes Eq. (1.34):

$$\mathcal{P}_>(\Lambda^*[n]) = n/N. \qquad (1.45)$$

[16] A third class of laws, stable under 'max' concerns random variables, which are bounded from above – i.e. such that $P(x) = 0$ for $x > x_M$, with x_M finite. This leads to the Weibull distributions, which we will not consider further in this book.

The width w_n of the distribution is found to be given by:

$$w_n = \frac{1}{\sqrt{N}} \frac{\sqrt{1-(n/N)^2}}{P(x = \Lambda^*[n])}, \qquad (1.46)$$

which shows that in the limit $N \to \infty$, the value of the nth variable is more and more sharply peaked around its most probable value $\Lambda^*[n]$, given by Eq. (1.45).

In the case of an exponential tail, one finds that $\Lambda^*[n] \simeq \log(N/n)/\alpha$; whereas in the case of power-law tails, one rather obtains:

$$\Lambda^*[n] \simeq A_+ \left(\frac{N}{n}\right)^{\frac{1}{\mu}}. \qquad (1.47)$$

This last equation shows that, for power-law variables, the encountered values are hierarchically organized: for example, the ratio of the largest value $x_{\max} \equiv \Lambda[1]$ to the second largest $\Lambda[2]$ is of the order of $2^{1/\mu}$, which becomes larger and larger as μ decreases, and conversely tends to one when $\mu \to \infty$.

The property, Eq. (1.47) is very useful in identifying empirically the nature of the tails of a probability distribution. One sorts in decreasing order the set of observed values $\{x_1, x_2, \ldots, x_N\}$ and one simply draws $\Lambda[n]$ as a function of n. If the variables are power-law distributed, this graph should be a straight line in log–log plot, with a slope $-1/\mu$, as given by Eq. (1.47) (Fig. 1.6). On the same figure, we have shown the result obtained for exponentially distributed variables. On this diagram, one observes an approximately straight line, but with an effective slope which varies with the total number of points N: the slope is less and less as N/n grows larger. In this sense, the formal remark made above, that an exponential distribution could be seen as a power-law with $\mu \to \infty$, becomes somewhat more concrete. Note that if the axes x and y of Figure 1.6 are interchanged, then according to Eq. (1.45), one obtains an estimate of the cumulative distribution, $\mathcal{P}_>$.

Let us finally note another property of power-laws, potentially interesting for their empirical determination. If one computes the average value of x conditioned to a certain minimum value Λ:

$$\langle x \rangle_\Lambda = \frac{\int_\Lambda^\infty x P(x)\,\mathrm{d}x}{\int_\Lambda^\infty P(x)\,\mathrm{d}x}, \qquad (1.48)$$

then, if $P(x)$ decreases as in Eq. (1.14), one finds, for $\Lambda \to \infty$,

$$\langle x \rangle_\Lambda = \frac{\mu}{\mu - 1}\Lambda, \qquad (1.49)$$

independently of the tail amplitude A_+^{μ}.[17] *The average $\langle x \rangle_\Lambda$ is thus always of the same order as Λ itself, with a proportionality factor which diverges as $\mu \to 1$.*

[17] This means that μ can be determined by a one parameter fit only.

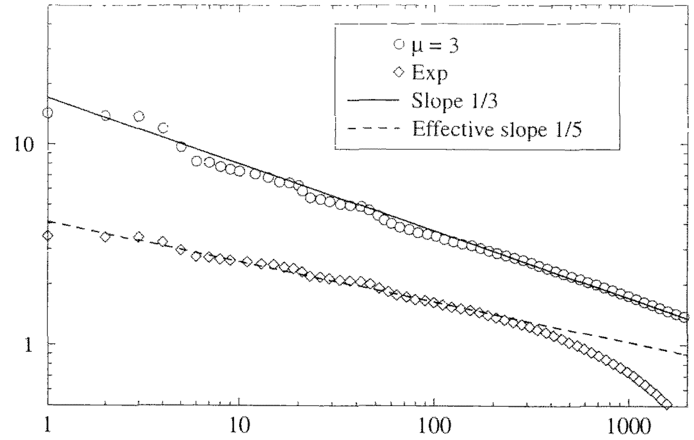

Fig. 1.6. Amplitude versus rank plots. One plots the value of the nth variable $\Lambda[n]$ as a function of its rank n. If $P(x)$ behaves asymptotically as a power-law, one obtains a straight line in log–log coordinates, with a slope equal to $-1/\mu$. For an exponential distribution, one observes an effective slope which is smaller and smaller as N/n tends to infinity. The points correspond to synthetic time series of length 5000, drawn according to a power-law with $\mu = 3$, or according to an exponential. Note that if the axes x and y are interchanged, then according to Eq. (1.45), one obtains an estimate of the cumulative distribution, $\mathcal{P}_>$.

1.5 Sums of random variables

In order to describe the statistics of future prices of a financial asset, one *a priori* needs a distribution density for all possible time intervals, corresponding to different trading time horizons. For example, the distribution of 5-min price fluctuations is different from the one describing daily fluctuations, itself different for the weekly, monthly, etc. variations. But in the case where the fluctuations are independent and identically distributed (iid), an assumption which is, however, usually not justified, see Sections 1.7 and 2.4, it is possible to reconstruct the distributions corresponding to different time scales from the knowledge of that describing short time scales only. In this context, Gaussians and Lévy distributions play a special role, because they are stable: if the short time scale distribution is a stable law, then the fluctuations on all time scales are described by the same stable law – only the parameters of the stable law must be changed (in particular its width). More generally, if one sums iid variables, then, independently of the short time distribution, the law describing long times converges towards one of the stable laws: this is the content of the 'central limit theorem' (CLT). In practice, however, this convergence can be very slow and thus of limited interest, in particular if one is concerned about short time scales.

1.5.1 Convolutions

What is the distribution of the sum of two independent random variable? This sum can, for example, represent the variation of price of an asset between today and the day after tomorrow (X), which is the sum of the increment between today and tomorrow (X_1) and between tomorrow and the day after tomorrow (X_2), both assumed to be random and independent.

Let us thus consider $X = X_1 + X_2$ where X_1 and X_2 are two random variables, independent, and distributed according to $P_1(x_1)$ and $P_2(x_2)$, respectively. The probability that X is equal to x (within dx) is given by the sum over all possibilities of obtaining $X = x$ (that is all combinations of $X_1 = x_1$ and $X_2 = x_2$ such that $x_1 + x_2 = x$), weighted by their respective probabilities. The variables X_1 and X_2 being independent, the joint probability that $X_1 = x_1$ and $X_2 = x - x_1$ is equal to $P_1(x_1)P_2(x - x_1)$, from which one obtains:

$$P(x, N = 2) = \int P_1(x')P_2(x - x')\,dx'. \tag{1.50}$$

This equation defines the convolution between $P_1(x)$ and $P_2(x)$, which we shall write $P = P_1 \star P_2$. The generalization to the sum of N independent random variables is immediate. If $X = X_1 + X_2 + \cdots + X_N$ with X_i distributed according to $P_i(x_i)$, the distribution of X is obtained as:

$$P(x, N) = \int P_1(x'_1)\ldots P_{N-1}(x'_{N-1})P_N(x - x'_1 - \cdots - x'_{N-1}) \prod_{i=1}^{N-1} dx'_i. \tag{1.51}$$

One thus understands how powerful is the hypothesis that the increments are iid, i.e. that $P_1 = P_2 = \cdots = P_N$. Indeed, according to this hypothesis, one only needs to know the distribution of increments over a unit time interval to reconstruct that of increments over an interval of length N: it is simply obtained by convoluting the elementary distribution N times with itself.

The analytical or numerical manipulations of Eqs (1.50) and (1.51) are much eased by the use of Fourier transforms, for which convolutions become simple products. The equation $P(x, N = 2) = [P_1 \star P_2](x)$, reads in Fourier space:

$$\hat{P}(z, N = 2) = \int e^{iz(x-x'+x')} \int P_1(x')P_2(x - x')\,dx'\,dx \equiv \hat{P}_1(z)\hat{P}_2(z). \tag{1.52}$$

In order to obtain the Nth convolution of a function with itself, one should raise its characteristic function to the power N, and then take its inverse Fourier transform.

1.5.2 Additivity of cumulants and of tail amplitudes

It is clear that the mean of the sum of two random variables (independent or not) is equal to the sum of the individual means. The mean is thus additive under

convolution. Similarly, if the random variables are independent, one can show that their variances (when they both exist) are also additive. More generally, all the cumulants (c_n) of two independent distributions simply add. This follows from the fact that since the characteristic functions multiply, their logarithm add. The additivity of cumulants is then a simple consequence of the linearity of derivation.

The cumulants of a given law convoluted N times with itself thus follow the simple rule $c_{n,N} = Nc_{n,1}$ where the $\{c_{n,1}\}$ are the cumulants of the elementary distribution P_1. Since the cumulant c_n has the dimension of X to the power n, its relative importance is best measured in terms of the normalized cumulants:

$$\lambda_n^N \equiv \frac{c_{n,N}}{(c_{2,N})^{\frac{n}{2}}} = \frac{c_{n,1}}{(c_{2,1})^{\frac{n}{2}}} N^{1-n/2}. \tag{1.53}$$

The normalized cumulants thus decay with N for $n > 2$; the higher the cumulant, the faster the decay: $\lambda_n^N \propto N^{1-n/2}$. The kurtosis κ, defined above as the fourth normalized cumulant, thus decreases as $1/N$. This is basically the content of the CLT: when N is very large, the cumulants of order > 2 become negligible. Therefore, the distribution of the sum is only characterized by its first two cumulants (mean and variance): it is a Gaussian.

Let us now turn to the case where the elementary distribution $P_1(x_1)$ decreases as a power-law for large arguments x_1 (cf. Eq. (1.14)), with a certain exponent μ. The cumulants of order higher than μ are thus divergent. By studying the small z singular expansion of the Fourier transform of $P(x, N)$, one finds that the above additivity property of cumulants is bequeathed to the tail amplitudes $A_\pm^\mu$: the asymptotic behaviour of the distribution of the sum $P(x, N)$ *still behaves as a power-law* (which is thus conserved by addition for all values of μ, provided one takes the limit $x \to \infty$ *before* $N \to \infty$ – see the discussion in Section 1.6.3), with a tail amplitude given by:

$$A_{\pm,N}^\mu \equiv NA_\pm^\mu. \tag{1.54}$$

The tail parameter thus plays the role, for power-law variables, of a generalized cumulant.

1.5.3 Stable distributions and self-similarity

If one adds random variables distributed according to an arbitrary law $P_1(x_1)$, one constructs a random variable which has, in general, a different probability distribution ($P(x, N) = [P_1(x_1)]^{\star N}$). However, for certain special distributions, the law of the sum has exactly the same shape as the elementary distribution – these are called *stable* laws. The fact that two distributions have the 'same shape' means that one can find a (N-dependent) translation and dilation of x such that the two laws coincide:

$$P(x, N)\,dx = P_1(x_1)\,dx_1 \quad \text{where } x = a_N x_1 + b_N. \tag{1.55}$$

The distribution of increments on a certain time scale (week, month, year) is thus *scale invariant*, provided the variable X is properly rescaled. In this case, the chart giving the evolution of the price of a financial asset as a function of time has the same statistical structure, independently of the chosen elementary time scale – only the average slope and the amplitude of the fluctuations are different. These charts are then called *self-similar*, or, using a better terminology introduced by Mandelbrot, *self-affine* (Figs 1.7 and 1.8).

The family of all possible stable laws coincide (for continuous variables) with the Lévy distributions defined above,[18] which include Gaussians as the special case $\mu = 2$. This is easily seen in Fourier space, using the explicit shape of the characteristic function of the Lévy distributions. We shall specialize here for simplicity to the case of symmetric distributions $P_1(x_1) = P_1(-x_1)$, for which the translation factor is zero ($b_N \equiv 0$). The scale parameter is then given by $a_N = N^{1/\mu}$,[19] and one finds, for $\mu < 2$:

$$\langle |x|^q \rangle^{\frac{1}{q}} \propto AN^{\frac{1}{\mu}} \qquad q < \mu \tag{1.56}$$

where $A = A_+ = A_-$. In words, the above equation means that the order of magnitude of the fluctuations on 'time' scale N is a factor $N^{1/\mu}$ larger than the fluctuations on the elementary time scale. However, once this factor is taken into account, the probability distributions are identical. One should notice the smaller the value of μ, the faster the growth of fluctuations with time.

1.6 Central limit theorem

We have thus seen that the stable laws (Gaussian and Lévy distributions) are 'fixed points' of the convolution operation. These fixed points are actually also *attractors*, in the sense that any distribution convoluted with itself a large number of times finally converges towards a stable law (apart from some very pathological cases). Said differently, the limit distribution of the sum of a large number of random variables is a stable law. The precise formulation of this result is known as the *central limit theorem* (CLT).

[18] For discrete variables, one should also add the Poisson distribution Eq. (1.27).
[19] The case $\mu = 1$ is special and involves extra logarithmic factors.

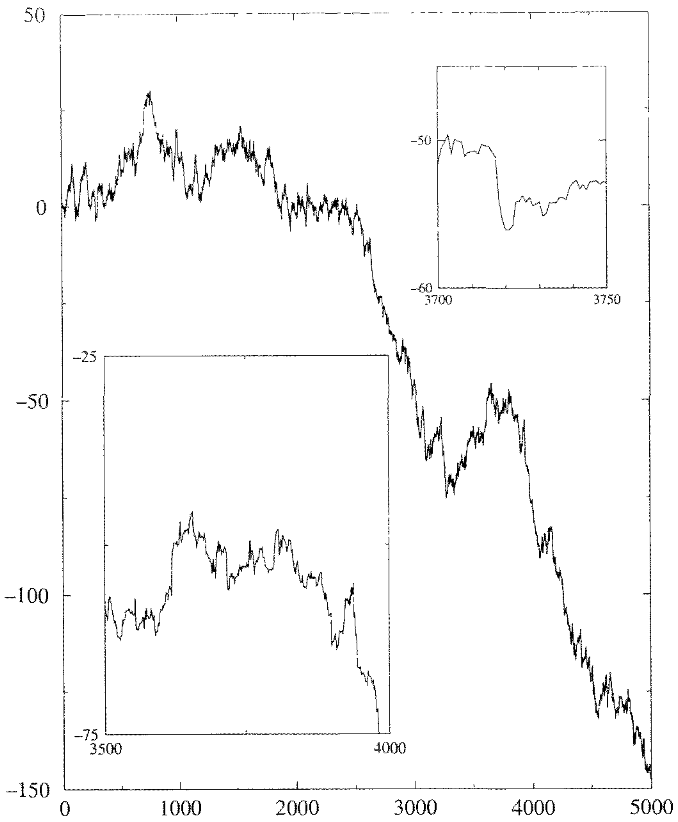

Fig. 1.7. Example of a self-affine function, obtained by summing random variables. One plots the sum x as a function of the number of terms N in the sum, for a Gaussian elementary distribution $P_1(x_1)$. Several successive 'zooms' reveal the self-similar nature of the function, here with $a_N = N^{1/2}$.

1.6.1 Convergence to a Gaussian

The classical formulation of the CLT deals with sums of iid random variables of *finite variance* σ^2 towards a Gaussian. In a more precise way, the result is then the following:

$$\lim_{N\to\infty} \mathcal{P}\left(u_1 \le \frac{x - mN}{\sigma\sqrt{N}} \le u_2\right) = \int_{u_1}^{u_2} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-u^2/2}\, \mathrm{d}u, \tag{1.57}$$

for all *finite* u_1, u_2. Note however that for finite N, the distribution of the sum $X = X_1 + \cdots + X_N$ in the tails (corresponding to extreme events) can be very different

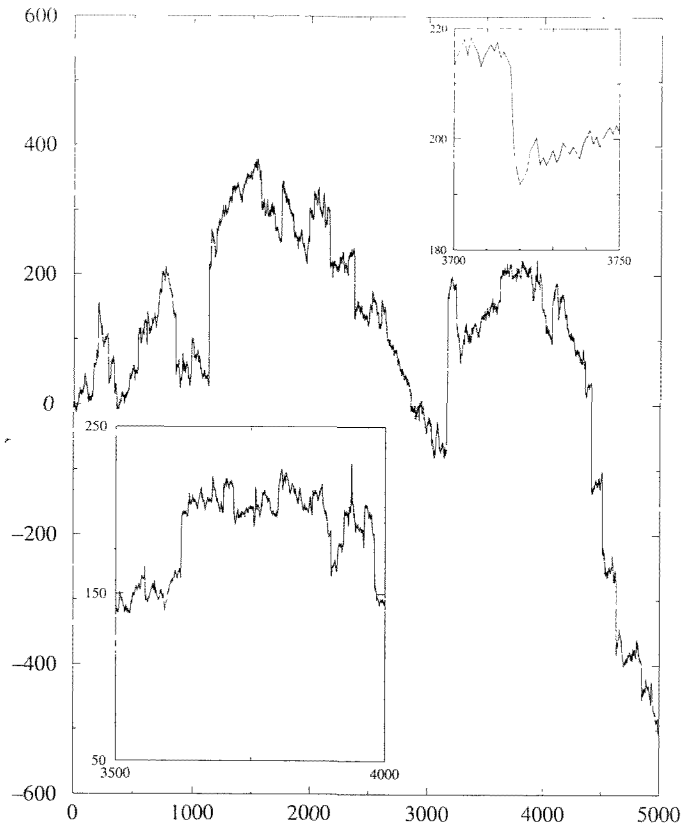

Fig. 1.8. In this case, the elementary distribution $P_1(x_1)$ decreases as a power-law with an exponent $\mu = 1.5$. The scale factor is now given by $a_N = N^{2/3}$. Note that, contrarily to the previous graph, one clearly observes the presence of sudden 'jumps', which reflect the existence of very large values of the elementary increment x_1.

from the Gaussian prediction; but the weight of these non-Gaussian regions tends to zero when N goes to infinity. The CLT only concerns the *central* region, which keeps a finite weight for N large: we shall come back in detail to this point below.

The main hypotheses ensuring the validity of the Gaussian CLT are the following:

- The X_i must be independent random variables, or at least not 'too' correlated (the correlation function $\langle x_i x_j \rangle - m^2$ must decay sufficiently fast when $|i - j|$ becomes large, see Section 1.7.1 below). For example, in the extreme case where all the X_i are perfectly correlated (i.e. they are all equal), the distribution of X

is obviously the same as that of the individual X_i (once the factor N has been properly taken into account).

- The random variables X_i need not necessarily be identically distributed. One must however require that the variance of all these distributions are not too dissimilar, so that no one of the variances dominates over all the others (as would be the case, for example, if the variances were themselves distributed as a power-law with an exponent $\mu < 1$). In this case, the variance of the Gaussian limit distribution is the average of the individual variances. This also allows one to deal with sums of the type $X = p_1 X_1 + p_2 X_2 + \cdots + p_N X_N$, where the p_i are arbitrary coefficients; this case is relevant in many circumstances, in particular in portfolio theory (cf. Chapter 3).
- Formally, the CLT only applies in the limit where N is infinite. In practice, N must be large enough for a Gaussian to be a good approximation of the distribution of the sum. The minimum required value of N (called N^* below) depends on the elementary distribution $P_1(x_1)$ and its distance from a Gaussian. Also, N^* depends on how far in the tails one requires a Gaussian to be a good approximation, which takes us to the next point.
- As mentioned above, the CLT does not tell us anything about the tails of the distribution of X; only the central part of the distribution is well described by a Gaussian. The 'central' region means a region of width at least of the order of $\sqrt{N}\sigma$ around the mean value of X. The actual width of the region where the Gaussian turns out to be a good approximation for large finite N crucially depends on the elementary distribution $P_1(x_1)$. This problem will be explored in Section 1.6.3. Roughly speaking, this region is of width $\sim N^{3/4}\sigma$ for 'narrow' symmetric elementary distributions, such that all even moments are finite. This region is however sometimes of much smaller extension: for example, if $P_1(x_1)$ has power-law tails with $\mu > 2$ (such that σ is finite), the Gaussian 'realm' grows barely faster than $\sqrt{N}$ (as $\sim \sqrt{N \log N}$).

The above formulation of the CLT requires the existence of a finite variance. This condition can be somewhat weakened to include some 'marginal' distributions such as a power-law with $\mu = 2$. In this case the scale factor is not $a_N = \sqrt{N}$ but rather $a_N = \sqrt{N \log N}$. However, as we shall discuss in the next section, elementary distributions which decay more slowly than $|x|^{-3}$ do not belong the the Gaussian basin of attraction. More precisely, the necessary and sufficient condition for $P_1(x_1)$ to belong to this basin is that:

$$\lim_{u\to\infty} u^2 \frac{\mathcal{P}_{1<}(-u) + \mathcal{P}_{1>}(u)}{\int_{|u'|<u} u'^2 P_1(u')\,\mathrm{d}u'} = 0. \tag{1.58}$$

This condition is always satisfied if the variance is finite, but allows one to include the marginal cases such as a power-law with $\mu = 2$.

The central limit theorem and information theory

It is interesting to notice that the Gaussian is the law of maximum entropy – or minimum information – such that its variance is fixed. The missing information quantity $\mathcal{I}$ (or entropy) associated with a probability distribution P is defined as:[20]

$$\mathcal{I}[P] \equiv -\int P(x)\log P(x)\,\mathrm{d}x. \tag{1.59}$$

The distribution maximizing $\mathcal{I}[P]$ for a given value of the variance is obtained by taking a functional derivative with respect to $P(x)$:

$$\frac{\partial}{\partial P(x)}\left[\mathcal{I}[P] - \zeta\int x'^2 P(x')\,\mathrm{d}x' - \zeta'\int P(x')\,\mathrm{d}x'\right] = 0, \tag{1.60}$$

where ζ is fixed by the condition $\int x^2 P(x)\,\mathrm{d}x = \sigma^2$ and ζ' by the normalization of $P(x)$. It is immediately seen that the solution to Eq. (1.60) is indeed the Gaussian. The numerical value of its entropy is:

$$\mathcal{I}_G = \frac{1}{2} + \frac{1}{2}\log(2\pi) + \log(\sigma) \simeq 1.419 + \log(\sigma). \tag{1.61}$$

For comparison, one can compute the entropy of the symmetric exponential distribution, which is:

$$\mathcal{I}_E = 1 + \frac{\log 2}{2} + \log(\sigma) \simeq 1.346 + \log(\sigma). \tag{1.62}$$

It is important to realize that the convolution operation is 'information burning', since all the details of the elementary distribution $P_1(x_1)$ progressively disappear while the Gaussian distribution emerges.

1.6.2 *Convergence to a Lévy distribution*

Let us now turn to the case of the sum of a large number N of iid random variables, asymptotically distributed as a power-law with $\mu < 2$, and with a tail amplitude $A^\mu = A_+^\mu = A_-^\mu$ (cf. Eq. (1.14)). The variance of the distribution is thus infinite. The limit distribution for large N is then a stable Lévy distribution of exponent μ and with a tail amplitude NA^μ. If the positive and negative tails of the elementary distribution $P_1(x_1)$ are characterized by different amplitudes (A_-^μ and A_+^μ) one then obtains an asymmetric Lévy distribution with parameter $\beta = (A_+^\mu - A_-^\mu)/(A_+^\mu + A_-^\mu)$. If the 'left' exponent is different from the 'right' exponent ($\mu_- \neq \mu_+$), then the smallest of the two wins and one finally obtains a totally asymmetric Lévy distribution ($\beta = -1$ or $\beta = 1$) with exponent $\mu = \min(\mu_-, \mu_+)$. The CLT generalized to Lévy distributions applies with the same precautions as in the Gaussian case above.

[20] Note that entropy is defined up to an additive constant. It is common to add 1 to the above definition.

Technically, a distribution $P_1(x_1)$ belongs to the basin of attraction of the Lévy distribution $L_{\mu,\beta}$ if and only if:

$$\lim_{u\to\infty} \frac{\mathcal{P}_{1<}(-u)}{\mathcal{P}_{1>}(u)} = \frac{1-\beta}{1+\beta}; \tag{1.63}$$

and for all r,

$$\lim_{u\to\infty} \frac{\mathcal{P}_{1<}(-u)+\mathcal{P}_{1>}(u)}{\mathcal{P}_{1<}(-ru)+\mathcal{P}_{1>}(ru)} = r^{\mu}. \tag{1.64}$$

A distribution with an asymptotic tail given by Eq. (1.14) is such that,

$$\mathcal{P}_{1<}(u) \underset{u\to-\infty}{\simeq} \frac{A_-^{\mu}}{|u|^{\mu}} \text{ and } \mathcal{P}_{1>}(u) \underset{u\to\infty}{\simeq} \frac{A_+^{\mu}}{u^{\mu}}, \tag{1.65}$$

and thus belongs to the attraction basin of the Lévy distribution of exponent μ and asymmetry parameter $\beta = (A_+^{\mu} - A_-^{\mu})/(A_+^{\mu} + A_-^{\mu})$.

1.6.3 Large deviations

The CLT teaches us that the Gaussian approximation is justified to describe the 'central' part of the distribution of the sum of a large number of random variables (of finite variance). However, the definition of the *centre* has remained rather vague up to now. The CLT only states that the probability of finding an event in the *tails* goes to zero for large N. In the present section, we characterize more precisely the region where the Gaussian approximation is valid.

If X is the sum of N iid random variables of mean m and variance σ^2, one defines a 'rescaled variable' U as:

$$U = \frac{X - Nm}{\sigma\sqrt{N}}, \tag{1.66}$$

which according to the CLT tends towards a Gaussian variable of zero mean and unit variance. Hence, for any *fixed* u, one has:

$$\lim_{N\to\infty} \mathcal{P}_>(u) = \mathcal{P}_{G>}(u), \tag{1.67}$$

where $\mathcal{P}_{G>}(u)$ is the related to the error function, and describes the weight contained in the tails of the Gaussian:

$$\mathcal{P}_{G>}(u) = \int_u^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-u'^2/2)\,\mathrm{d}u' = \tfrac{1}{2}\mathrm{erfc}\left(\frac{u}{\sqrt{2}}\right). \tag{1.68}$$

However, the above convergence is *not uniform*. The value of N such that the approximation $\mathcal{P}_>(u) \simeq \mathcal{P}_{G>}(u)$ becomes valid depends on u. Conversely, for fixed N, this approximation is only valid for u not too large: $|u| \ll u_0(N)$.

One can estimate $u_0(N)$ in the case where the elementary distribution $P_1(x_1)$ is 'narrow', that is, decreasing faster than any power-law when $|x_1| \to \infty$, such that all the moments are finite. In this case, all the cumulants of P_1 are finite and one can obtain a systematic expansion in powers of $N^{-1/2}$ of the difference $\Delta\mathcal{P}_>(u) \equiv \mathcal{P}_>(u) - \mathcal{P}_{G>}(u)$,

$$\Delta\mathcal{P}_>(u) \simeq \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \left(\frac{Q_1(u)}{N^{1/2}} + \frac{Q_2(u)}{N} + \cdots + \frac{Q_k(u)}{N^{k/2}} + \cdots\right), \tag{1.69}$$

where the $Q_k(u)$ are polynomials functions which can be expressed in terms of the normalized cumulants λ_n (cf. Eq. (1.12)) of the elementary distribution. More explicitly, the first two terms are given by:

$$Q_1(u) = \tfrac{1}{6}\lambda_3(u^2 - 1), \tag{1.70}$$

and

$$Q_2(u) = \tfrac{1}{72}\lambda_3^2 u^5 + \tfrac{1}{8}\left(\tfrac{1}{3}\lambda_4 - \tfrac{10}{9}\lambda_3^2\right)u^3 + \left(\tfrac{5}{24}\lambda_3^2 - \tfrac{1}{8}\lambda_4\right)u. \tag{1.71}$$

One recovers the fact that if all the cumulants of $P_1(x_1)$ of order larger than two are zero, all the Q_k are also identically zero and so is the difference between $P(x, N)$ and the Gaussian.

For a general asymmetric elementary distribution P_1, λ_3 is non-zero. The leading term in the above expansion when N is large is thus $Q_1(u)$. For the Gaussian approximation to be meaningful, one must at least require that this term is small in the central region where u is of order one, which corresponds to $x - mN \sim \sigma\sqrt{N}$. This thus imposes that $N \gg N^* = \lambda_3^2$. The Gaussian approximation remains valid whenever the relative error is small compared to 1. For large u (which will be justified for large N), the relative error is obtained by dividing Eq. (1.69) by $\mathcal{P}_{G>}(u) \simeq \exp(-u^2/2)/(u\sqrt{2\pi})$. One then obtains the following condition:[21]

$$\lambda_3 u^3 \ll N^{1/2} \quad \text{i.e. } |x - Nm| \ll \sigma\sqrt{N}\left(\frac{N}{N^*}\right)^{1/6}. \tag{1.72}$$

This shows that the central region has an extension growing as $N^{2/3}$.

A symmetric elementary distribution is such that $\lambda_3 \equiv 0$; it is then the kurtosis $\kappa = \lambda_4$ that fixes the first correction to the Gaussian when N is large, and thus the extension of the central region. The conditions now read: $N \gg N^* = \lambda_4$ and

$$\lambda_4 u^4 \ll N \quad \text{i.e. } |x - Nm| \ll \sigma\sqrt{N}\left(\frac{N}{N^*}\right)^{1/4}. \tag{1.73}$$

The central region now extends over a region of width $N^{3/4}$.

The results of the present section do not directly apply if the elementary distribution $P_1(x_1)$ decreases as a power-law ('broad distribution'). In this case, some of the cumulants are infinite and the above cumulant expansion, Eq. (1.69), is

[21] The above arguments can actually be made fully rigorous, see [Feller].

meaningless. In the next section, we shall see that in this case the 'central' region is much more restricted than in the case of 'narrow' distributions. We shall then describe in Section 1.6.5, the case of 'truncated' power-law distributions, where the above conditions become asymptotically relevant. These laws however may have a very large kurtosis, which depends on the point where the truncation becomes noticeable, and the above condition $N \gg \lambda_4$ can be hard to satisfy.

Cramèr function

More generally, when N is large, one can write the distribution of the sum of N iid random variables as:[22]

$$P(x, N) \underset{N\to\infty}{\simeq} \exp\left[-NS\left(\frac{x}{N}\right)\right], \tag{1.74}$$

where S is the so-called Cramèr function, which gives some information about the probability of X even outside the 'central' region. When the variance is finite, S grows as $S(u) \propto u^2$ for small u's, which again leads to a Gaussian central region. For finite u, S can be computed using Laplace's saddle point method, valid for N large. By definition:

$$P(x, N) = \frac{1}{2\pi}\int \exp N\left(-iz\frac{x}{N} + \log[\hat{P}_1(z)]\right) dz. \tag{1.75}$$

When N is large, the above integral is dominated by the neighbourhood of the point z^* where the term in the exponential is stationary. The results can be written as:

$$P(x, N) \simeq \exp\left[-NS\left(\frac{x}{N}\right)\right], \tag{1.76}$$

with $S(u)$ given by:

$$\left.\frac{d\log[\hat{P}_1(z)]}{dz}\right|_{z=z^*} = iu \qquad S(u) = -iz^*u + \log[\hat{P}_1(z^*)]. \tag{1.77}$$

which, in principle, allows one to estimate $P(x, N)$ even outside the central region. Note that if $S(u)$ is finite for finite u, the corresponding probability is exponentially small in N.

1.6.4 The CLT at work on a simple case

It is helpful to give some flesh to the above general statements, by working out explicitly the convergence towards the Gaussian in two exactly soluble cases. On these examples, one clearly sees the domain of validity of the CLT as well as its limitations.

Let us first study the case of positive random variables distributed according to the exponential distribution:

$$P_1(x) = \Theta(x_1)\alpha e^{-\alpha x_1}, \tag{1.78}$$

where $\Theta(x_1)$ is the function equal to 1 for $x_1 \geq 0$ and to 0 otherwise. A simple

[22] We assume that their mean is zero, which can always be achieved through a suitable shift of x_1.

computation shows that the above distribution is correctly normalized, has a mean given by $m = \alpha^{-1}$ and a variance given by $\sigma^2 = \alpha^{-2}$. Furthermore, the exponential distribution is asymmetrical; its skewness is given by $c_3 = \langle (x-m)^3 \rangle = 2\alpha^{-3}$, or $\lambda_3 = 2$.

The sum of N such variables is distributed according to the Nth convolution of the exponential distribution. According to the CLT this distribution should approach a Gaussian of mean mN and of variance $N\sigma^2$. The Nth convolution of the exponential distribution can be computed exactly. The result is:[23]

$$P(x, N) = \Theta(x)\alpha^N \frac{x^{N-1}e^{-\alpha x}}{(N-1)!}, \tag{1.79}$$

which is called a 'Gamma' distribution of index N. At first sight, this distribution does not look very much like a Gaussian! For example, its asymptotic behaviour is very far from that of a Gaussian: the 'left' side is strictly zero for negative x, while the 'right' tail is exponential, and thus much fatter than the Gaussian. It is thus very clear that the CLT does not apply for values of x too far from the mean value. However, the central region around $Nm = N\alpha^{-1}$ is well described by a Gaussian. The most probable value (x^*) is defined as:

$$\left.\frac{d}{dx}x^{N-1}e^{-\alpha x}\right|_{x^*} = 0, \tag{1.80}$$

or $x^* = (N-1)m$. An expansion in $x - x^*$ of $P(x, N)$ then gives us:

$$\begin{aligned}\log P(x, N) &= -K(N-1) - \log m - \frac{\alpha^2(x-x^*)^2}{2(N-1)} \\ &\quad + \frac{\alpha^3(x-x^*)^3}{3(N-1)^2} + O(x-x^*)^4,\end{aligned} \tag{1.81}$$

where

$$K(N) \equiv \log N! + N - N\log N \underset{N\to\infty}{\simeq} \tfrac{1}{2}\log(2\pi N). \tag{1.82}$$

Hence, to second order in $x - x^*$, $P(x, N)$ is given by a Gaussian of mean $(N-1)m$ and variance $(N-1)\sigma^2$. The relative difference between N and $N-1$ goes to zero for large N. Hence, for the Gaussian approximation to be valid, one requires not only that N be large compared to one, but also that the higher-order terms in $(x - x^*)$ be negligible. The cubic correction is small compared to 1 as long as $\alpha|x - x^*| \ll N^{2/3}$, in agreement with the above general statement, Eq. (1.72), for an elementary distribution with a non-zero third cumulant. Note also that for $x \to \infty$, the exponential behaviour of the Gamma function coincides (up

[23] This result can be shown by induction using the definition (Eq. (1.50)).

to subleading terms in x^{N-1}) with the asymptotic behaviour of the elementary distribution $P_1(x_1)$.

Another very instructive example is provided by a distribution which behaves as a power-law for large arguments, but at the same time has a finite variance to ensure the validity of the CLT. Consider the following explicit example of a Student distribution with $\mu = 3$:

$$P_1(x_1) = \frac{2a^3}{\pi(x_1^2 + a^2)^2}, \tag{1.83}$$

where a is a positive constant. This symmetric distribution behaves as a power-law with $\mu = 3$ (cf. Eq. (1.14)); all its cumulants of order larger than or equal to three are infinite. However, its variance is finite and equal to a^2.

It is useful to compute the characteristic function of this distribution,

$$\hat{P}_1(z) = (1 + a|z|)e^{-a|z|}, \tag{1.84}$$

and the first terms of its small z expansion, which read:

$$\hat{P}_1(z) \simeq 1 - \frac{z^2 a^2}{2} + \frac{|z|^3 a^3}{3} + O(z^4). \tag{1.85}$$

The first singular term in this expansion is thus $|z|^3$, as expected from the asymptotic behaviour of $P_1(x_1)$ in x_1^{-4}, and the divergence of the moments of order larger than three.

The Nth convolution of $P_1(x_1)$ thus has the following characteristic function:

$$\hat{P}_1^{\,N}(z) = (1 + a|z|)^N e^{-aN|z|}, \tag{1.86}$$

which, expanded around $z = 0$, gives:

$$\hat{P}_1^{\,N}(k) \simeq 1 - \frac{Nz^2 a^2}{2} + \frac{N|z|^3 a^3}{3} + O(z^4). \tag{1.87}$$

Note that the $|z|^3$ singularity (which signals the divergence of the moments m_n for $n \geq 3$) does not disappear under convolution, even if at the same time $P(x, N)$ converges towards the Gaussian. The resolution of this apparent paradox is again that the convergence towards the Gaussian only concerns the centre of the distribution, whereas the tail in x^{-4} survives for ever (as was mentioned in Section 1.5.3).

As follows from the CLT, the centre of $P(x, N)$ is well approximated, for N large, by a Gaussian of zero mean and variance Na^2:

$$P(x, N) \simeq \frac{1}{\sqrt{2\pi N}a} \exp\left(-\frac{x^2}{2Na^2}\right). \tag{1.88}$$

On the other hand, since the power-law behaviour is conserved upon addition and that the tail amplitudes simply add (cf. Eq. (1.14)), one also has, for large x's:

$$P(x, N) \underset{x\to\infty}{\simeq} \frac{2Na^3}{\pi x^4}. \tag{1.89}$$

The above two expressions Eqs (1.88) and (1.89) are not incompatible, since these describe two very different regions of the distribution $P(x, N)$. For fixed N, there is a characteristic value $x_0(N)$ beyond which the Gaussian approximation for $P(x, N)$ is no longer accurate, and the distribution is described by its asymptotic power-law regime. The order of magnitude of $x_0(N)$ is fixed by looking at the point where the two regimes match to one another:

$$\frac{1}{\sqrt{2\pi N}a} \exp\left(-\frac{x_0^2}{2Na^2}\right) \simeq \frac{2Na^3}{\pi x_0^4}. \tag{1.90}$$

One thus finds,

$$x_0(N) \simeq a\sqrt{N \log N}, \tag{1.91}$$

(neglecting subleading corrections for large N).

This means that the rescaled variable $U = X/(a\sqrt{N})$ becomes for large N a Gaussian variable of unit variance, but this description ceases to be valid as soon as $u \sim \sqrt{\log N}$, which grows very slowly with N. For example, for N equal to a million, the Gaussian approximation is only acceptable for fluctuations of u of less than three or four RMS!

Finally, the CLT states that the weight of the regions where $P(x, N)$ substantially differs from the Gaussian goes to zero when N becomes large. For our example, one finds that the probability that X falls in the tail region rather than in the central region is given by:

$$\mathcal{P}_<(x_0) + \mathcal{P}_>(x_0) \simeq 2\int_{a\sqrt{N \log N}}^{\infty} \frac{2a^3 N}{\pi x^4}\,dx \propto \frac{1}{\sqrt{N}\log^{3/2} N}, \tag{1.92}$$

which indeed goes to zero for large N.

The above arguments are not special to the case $\mu = 3$ and in fact apply more generally, as long as $\mu > 2$, i.e. when the variance is finite. In the general case, one finds that the CLT is valid in the region $|x| \ll x_0 \propto \sqrt{N \log N}$, and that the weight of the non-Gaussian tails is given by:

$$\mathcal{P}_<(x_0) + \mathcal{P}_>(x_0) \propto \frac{1}{N^{\mu/2-1}\log^{\mu/2} N}, \tag{1.93}$$

which tends to zero for large N. However, one should notice that as μ approaches the 'dangerous' value $\mu = 2$, the weight of the tails becomes more and more important. For $\mu < 2$, the whole argument collapses since the weight of the tails would grow with N. In this case, however, the convergence is no longer towards the Gaussian, but towards the Lévy distribution of exponent μ.

1.6.5 Truncated Lévy distributions

An interesting case is when the elementary distribution $P_1(x_1)$ is a truncated Lévy distribution (TLD) as defined in Section 1.3.3. The first cumulants of the distribution defined by Eq. (1.23) read, for $1 < \mu < 2$:

$$c_2 = \mu(\mu-1)\frac{a_\mu}{|\cos\pi\mu/2|}\alpha^{\mu-2} \qquad c_3 = 0. \tag{1.94}$$

The kurtosis $\kappa = \lambda_4 = c_4/c_2^2$ is given by:

$$\lambda_4 = \frac{(3-\mu)(2-\mu)|\cos\pi\mu/2|}{\mu(\mu-1)a_\mu\alpha^\mu}. \tag{1.95}$$

Note that the case $\mu = 2$ corresponds to the Gaussian, for which $\lambda_4 = 0$ as expected. On the other hand, when $\alpha \to 0$, one recovers a pure Lévy distribution, for which c_2 and c_4 are formally infinite. Finally, if $\alpha \to \infty$ with $a_\mu\alpha^{\mu-2}$ fixed, one also recovers the Gaussian.

If one considers the sum of N random variables distributed according to a TLD, the condition for the CLT to be valid reads (for $\mu < 2$):[24]

$$N \gg N^* = \lambda_4 \Longrightarrow (Na_\mu)^{\frac{1}{\mu}} \gg \alpha^{-1}. \tag{1.96}$$

This condition has a very simple intuitive meaning. A TLD behaves very much like a pure Lévy distribution as long as $x \ll \alpha^{-1}$. In particular, it behaves as a power-law of exponent μ and tail amplitude $A^\mu \propto a_\mu$ in the region where x is large but still much smaller than α^{-1} (we thus also assume that α is very small). If N is not too large, most values of x fall in the Lévy-like region. The largest value of x encountered is thus of order $x_{\max} \simeq AN^{1/\mu}$ (cf. Eq. (1.42)). If $x_{\max}$ is very small compared to α^{-1}, it is consistent to forget the exponential cut-off and think of the elementary distribution as a pure Lévy distribution. One thus observe a first regime in N where the typical value of X grows as $N^{1/\mu}$, as if α was zero.[25] However, as illustrated in Figure 1.9, this regime ends when $x_{\max}$ reaches the cut-off value α^{-1}: this happens precisely when N is of the order of N^* defined above. For $N > N^*$, the variable X progressively converges towards a Gaussian variable of width $\sqrt{N}$, at least in the region where $|x| \ll \sigma N^{3/4}/N^{*1/4}$. The typical amplitude of X thus behaves (as a function of N) as sketched in Figure 1.9. Notice that the asymptotic part of the distribution of X (outside the central region) decays as an exponential for all values of N.

[24] One can see by inspection that the other conditions, concerning higher-order cumulants, and which read $N^{k-1}\lambda_{2k} \gg 1$, are actually equivalent to the one written here.

[25] Note however that the variance of X grows like N for all N. However, the variance is dominated by the cut-off and, in the region $N \ll N^*$, grossly overestimates the typical values of X, see Section 2.3.2.

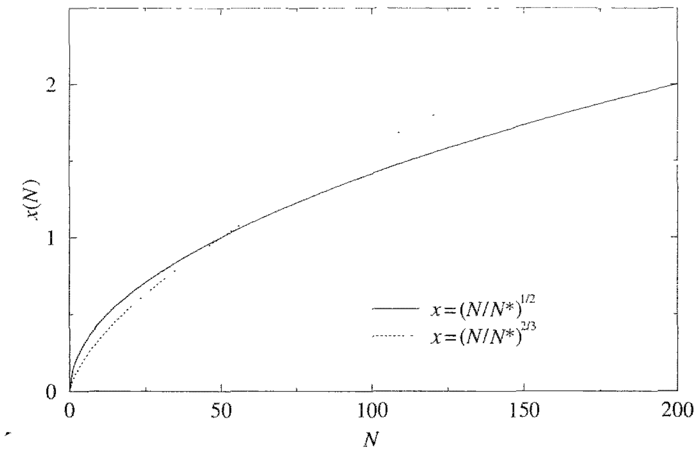

Fig. 1.9. Behaviour of the typical value of X as a function of N for TLD variables. When $N \ll N^*$, x grows as $N^{1/\mu}$ (dotted line). When $N \sim N^*$, x reaches the value α^{-1} and the exponential cut-off starts being relevant. When $N \gg N^*$, the behaviour predicted by the CLT sets in, and one recovers $x \propto \sqrt{N}$ (plain line).

1.6.6 Conclusion: survival and vanishing of tails

The CLT thus teaches us that if the number of terms in a sum is large, the sum becomes (nearly) a Gaussian variable. This sum can represent the temporal aggregation of the daily fluctuations of a financial asset, or the aggregation, in a portfolio, of different stocks. The Gaussian (or non-Gaussian) nature of this sum is thus of crucial importance for risk control, since the extreme tails of the distribution correspond to the most 'dangerous' fluctuations. As we have discussed above, fluctuations are never Gaussian in the far-tails: one can explicitly show that if the elementary distribution decays as a power-law (or as an exponential, which formally corresponds to $\mu = \infty$), the distribution of the sum decays in the very same manner outside the central region, i.e. much more slowly than the Gaussian. The CLT simply ensures that these tail regions are expelled more and more towards large values of X when N grows, and their associated probability is smaller and smaller. When confronted with a concrete problem, one must decide whether N is large enough to be satisfied with a Gaussian description of the risks. In particular, if N is less than the characteristic value N^* defined above, the Gaussian approximation is very bad.

1.7 Correlations, dependence and non-stationary models (*)

We have assumed up to now that the random variables were *independent* and *identically distributed*. Although the general case cannot be discussed as thoroughly as the iid case, it is useful to illustrate how the CLT must be modified on a few examples, some of which are particularly relevant in the context of financial time series.

1.7.1 Correlations

Let us assume that the correlation function $C_{i,j}$ (defined as $\langle x_i x_j \rangle - m^2$) of the random variables is non-zero for $i \neq j$. We also assume that the process is *stationary*, i.e. that $C_{i,j}$ only depends on $|i-j|$: $C_{i,j} = C(|i-j|)$, with $C(\infty) = 0$. The variance of the sum can be expressed in terms of the matrix C as:[26]

$$\langle x^2 \rangle = \sum_{i,j=1}^{N} C_{i,j} = N\sigma^2 + 2N \sum_{\ell=1}^{N} \left(1 - \frac{\ell}{N}\right) C(\ell), \tag{1.97}$$

where $\sigma^2 \equiv C(0)$. From this expression, it is readily seen that if $C(\ell)$ decays faster than $1/\ell$ for large ℓ, the sum over ℓ tends to a constant for large N, and thus the variance of the sum still grows as N, as for the usual CLT. If however $C(\ell)$ decays for large ℓ as a power-law $\ell^{-\nu}$, with $\nu < 1$, then the variance grows faster than N, as $N^{2-\nu}$ – correlations thus enhance fluctuations. Hence, when $\nu < 1$, the standard CLT certainly has to be amended. The problem of the limit distribution in these cases is however not solved in general. For example, if the X_i are correlated Gaussian variables, it is easy to show that the resulting sum is also Gaussian, whatever the value of ν. Another solvable case is when the X_i are correlated Gaussian variables, but one takes the sum of the *squares* of the X_i's. This sum converges towards a Gaussian of width $\sqrt{N}$ whenever $\nu > \frac{1}{2}$, but towards a non-trivial limit distribution of a new kind (i.e. neither Gaussian nor Lévy stable) when $\nu < \frac{1}{2}$. In this last case, the proper rescaling factor must be chosen as $N^{1-\nu}$.

One can also construct *anti-correlated* random variables, the sum of which grows slower than $\sqrt{N}$. In the case of power-law correlated or anti-correlated Gaussian random variables, one speaks of 'fractional Brownian motion'. This notion was introduced in [Mandelbrot and Van Ness].

1.7.2 Non-stationary models and dependence

It may happen that the distribution of the elementary random variables $P_1(x_1)$, $P_2(x_2), \ldots, P_N(x_N)$ are not all identical. This is the case, for example, when the

[26] We again assume in the following, without loss of generality, that the mean m is zero.

variance of the random process depends upon time – in financial markets, it is a well-known fact that the daily volatility is time dependent, taking rather high levels in periods of uncertainty, and reverting back to lower values in calmer periods. For example, the volatility of the bond market has been very high during 1994, and decreased in later years. Similarly, the volatility of stock markets has increased since August 1997.

If the distribution P_k varies sufficiently 'slowly', one can in principle measure some of its moments (for example its mean and variance) over a time scale which is long enough to allow for a precise determination of these moments, but short compared to the time scale over which P_k is expected to vary. The situation is less clear if P_k varies 'rapidly'. Suppose for example that $P_k(x_k)$ is a Gaussian distribution of variance σ_k^2, which is itself a random variable. We shall denote as $\overline{(\cdots)}$ the average over the random variable σ_k, to distinguish it from the notation $\langle \cdots \rangle_k$ which we have used to describe the average over the probability distribution P_k. If σ_k varies rapidly, it is impossible to separate the two sources of uncertainty. Thus, the empirical histogram constructed from the series $\{x_1, x_2, \ldots, x_N\}$ leads to an 'apparent' distribution $\overline{P}$ which is non-Gaussian even if each individual P_k is Gaussian. Indeed, from:

$$\overline{P}(x) \equiv \int P(\sigma) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) d\sigma, \tag{1.98}$$

one can calculate the kurtosis of $\overline{P}$ as:

$$\overline{\kappa} = \frac{\overline{\langle x^4 \rangle}}{(\overline{\langle x^2 \rangle})^2} - 3 \equiv 3\left(\frac{\overline{\sigma^4}}{(\overline{\sigma^2})^2} - 1\right). \tag{1.99}$$

Since for any random variable one has $\overline{\sigma^4} \geq (\overline{\sigma^2})^2$ (the equality being reached only if σ^2 does not fluctuate at all), one finds that $\overline{\kappa}$ is always positive. The volatility fluctuations can thus lead to 'fat tails'. More precisely, let us assume that the probability distribution of the RMS, $P(\sigma)$, decays itself for large σ as $\exp(-\sigma^c)$, $c > 0$. Assuming P_k to be Gaussian, it is easy to obtain, using a saddle-point method (cf. Eq. (1.75)), that for large x one has:

$$\log[\overline{P}(x)] \propto -x^{\frac{2c}{2+c}}. \tag{1.100}$$

Since $c < 2 + c$, this asymptotic decay is always much slower than in the Gaussian case, which corresponds to $c \to \infty$. The case where the volatility itself has a Gaussian tail ($c = 2$) leads to an exponential decay of $\overline{P}(x)$.

Another interesting case is when σ^2 is distributed as a completely asymmetric Lévy distribution ($\beta = 1$) of exponent $\mu < 1$. Using the properties of Lévy distributions, one can then show that $\overline{P}$ is itself a symmetric Lévy distribution ($\beta = 0$), of exponent equal to 2μ.

If the fluctuations of σ_k are themselves correlated, one observes an interesting case of *dependence*. For example, if σ_k is large, σ_{k+1} will probably also be large. The fluctuation X_k thus has a large probability to be large (but of arbitrary sign) twice in a row. We shall often refer, in the following, to a simple model where x_k can be written as a product $\epsilon_k\sigma_k$, where ϵ_k are iid random variables of zero mean and unit variance, and σ_k corresponds to the local 'scale' of the fluctuations, which can be correlated in time. The correlation function of the X_k is thus given by:

$$\overline{\langle x_i x_j\rangle} = \overline{\sigma_i\sigma_j}\langle\epsilon_i\epsilon_j\rangle = \delta_{i,j}\overline{\sigma^2}. \tag{1.101}$$

Hence the X_k are uncorrelated random variables, but they are not independent since a higher-order correlation function reveals a richer structure. Let us for example consider the correlation of X_k^2:

$$\overline{\langle x_i^2 x_j^2\rangle} - \overline{\langle x_i^2\rangle\langle x_j^2\rangle} = \overline{\sigma_i^2\sigma_j^2} - \overline{\sigma_i^2}\,\overline{\sigma_j^2} \qquad (i \neq j), \tag{1.102}$$

which indeed has an interesting temporal behaviour: see Section 2.4.[27] However, even if the correlation function $\overline{\sigma_i^2\sigma_j^2} - \overline{\sigma^2}^2$ decreases very slowly with $|i-j|$, one can show that the sum of the X_k, obtained as $\sum_{k=1}^N \epsilon_k\sigma_k$ is still governed by the CLT, and converges for large N towards a Gaussian variable. A way to see this is to compute the average kurtosis of the sum, κ_N. As shown in Appendix A, one finds the following result:

$$\kappa_N = \frac{1}{N}\left[\kappa_0 + (3+\kappa_0)g(0) + 6\sum_{\ell=1}^{N}\left(1-\frac{\ell}{N}\right)g(\ell)\right], \tag{1.103}$$

where κ_0 is the kurtosis of the variable ϵ, and $g(\ell)$ the correlation function of the variance, defined as:

$$\overline{\sigma_i^2\sigma_j^2} - \overline{\sigma^2}^2 = \overline{\sigma^2}^2 g(|i-j|). \tag{1.104}$$

It is interesting to see that for $N = 1$, the above formula gives $\kappa_1 = \kappa_0 + (3+\kappa_0)g(0) > \kappa_0$, which means that even if $\kappa_0 = 0$, a fluctuating volatility is enough to produce some kurtosis. More importantly, one sees that if the variance correlation function $g(\ell)$ decays with ℓ, the kurtosis κ_N tends to zero with N, thus showing that the sum indeed converges towards a Gaussian variable. For example, if $g(\ell)$ decays as a power-law $\ell^{-\nu}$ for large ℓ, one finds that for large N:

$$\kappa_N \propto \frac{1}{N} \quad \text{for} \quad \nu > 1; \qquad \kappa_N \propto \frac{1}{N^\nu} \quad \text{for} \quad \nu < 1. \tag{1.105}$$

[27] Note that for $i \neq j$ this correlation function can be zero either because σ is identically equal to a certain value σ_0, or because the fluctuations of σ are completely uncorrelated from one time to the next.

Hence, long-range correlation in the variance considerably slows down the convergence towards the Gaussian. This remark will be of importance in the following, since financial time series often reveal long-ranged volatility fluctuations.

1.8 Central limit theorem for random matrices (*)

One interesting application of the CLT concerns the spectral properties of 'random matrices'. The theory of random matrices has made enormous progress during the past 30 years, with many applications in physical sciences and elsewhere. More recently, it has been suggested that random matrices might also play an important role in finance: an example is discussed in Section 2.7. It is therefore appropriate to give a cursory discussion of some salient properties of random matrices. The simplest ensemble of random matrices is one where all elements of the matrix **H** are iid random variables, with the only constraint that the matrix be symmetrical ($H_{ij} = H_{ji}$). One interesting result is that in the limit of very large matrices, the distribution of its eigenvalues has universal properties, which are to a large extent independent of the distribution of the elements of the matrix. This is actually the consequence of the CLT, as we will show below. Let us introduce first some notation. The matrix **H** is a square, $M \times M$ symmetric matrix. Its eigenvalues are λ_α, with $\alpha = 1, \ldots, M$. The *density* of eigenvalues is defined as:

$$\rho(\lambda) = \frac{1}{M}\sum_{\alpha=1}^{M}\delta(\lambda-\lambda_\alpha), \tag{1.106}$$

where δ is the Dirac function. We shall also need the so-called 'resolvent' $\mathbf{G}(\lambda)$ of the matrix **H**, defined as:

$$G_{ij}(\lambda) \equiv \left(\frac{1}{\lambda\mathbf{1}-\mathbf{H}}\right)_{ij}, \tag{1.107}$$

where $\mathbf{1}$ is the identity matrix. The trace of $\mathbf{G}(\lambda)$ can be expressed using the eigenvalues of **H** as:

$$\mathrm{Tr}\,\mathbf{G}(\lambda) = \sum_{\alpha=1}^{M}\frac{1}{\lambda-\lambda_\alpha}. \tag{1.108}$$

The 'trick' that allows one to calculate $\rho(\lambda)$ in the large M limit is the following representation of the δ function:

$$\frac{1}{x-\mathrm{i}\epsilon} = PP\frac{1}{x} + \mathrm{i}\pi\delta(x) \qquad (\epsilon \to 0), \tag{1.109}$$

where PP means the principal part. Therefore, $\rho(\lambda)$ can be expressed as:

$$\rho(\lambda) = \lim_{\epsilon\to 0}\frac{1}{M\pi}\Im\left(\mathrm{Tr}\,\mathbf{G}(\lambda-\mathrm{i}\epsilon)\right). \tag{1.110}$$

Our task is therefore to obtain an expression for the resolvent $\mathbf{G}(\lambda)$. This can be done by establishing a recursion relation, allowing one to compute $\mathbf{G}(\lambda)$ for a matrix $\mathbf{H}$ with one extra row and one extra column, the elements of which being H_{0i}. One then computes $G_{00}^{M+1}(\lambda)$ (the superscript stands for the size of the matrix $\mathbf{H}$) using the standard formula for matrix inversion:

$$G_{00}^{M+1}(\lambda) = \frac{\text{minor}(\lambda\mathbf{1} - \mathbf{H})_{00}}{\det(\lambda\mathbf{1} - \mathbf{H})}. \tag{1.111}$$

Now, one expands the determinant appearing in the denominator in minors along the first row, and then each minor is itself expanded in subminors along their first column. After a little thought, this finally leads to the following expression for $G_{00}^{M+1}(\lambda)$:

$$\frac{1}{G_{00}^{M+1}(\lambda)} = \lambda - H_{00} - \sum_{i,j=1}^{M} H_{0i} H_{0j} G_{ij}^{M}(\lambda). \tag{1.112}$$

This relation is general, without any assumption on the H_{ij}. Now, we assume that the H_{ij}'s are iid random variables, of zero mean and variance equal to $\langle H_{ij}^2 \rangle = \sigma^2/M$. This scaling with M can be understood as follows: when the matrix H acts on a certain vector, each component of the image vector is a sum of M random variables. In order to keep the image vector (and thus the corresponding eigenvalue) finite when $M \to \infty$, one should scale the elements of the matrix with the factor $1/\sqrt{M}$.

One could also write a recursion relation for G_{0i}^{M+1}, and establish self-consistently that $G_{ij} \sim 1/\sqrt{M}$ for $i \neq j$. On the other hand, due to the diagonal term λ, G_{ii} remains finite for $M \to \infty$. This scaling allows us to discard all the terms with $i \neq j$ in the sum appearing in the right-hand side of Eq. (1.112). Furthermore, since $H_{00} \sim 1/\sqrt{M}$, this term can be neglected compared to λ. This finally leads to a simplified recursion relation, valid in the limit $M \to \infty$:

$$\frac{1}{G_{00}^{M+1}(\lambda)} \simeq \lambda - \sum_{i=1}^{M} H_{0i}^2 G_{ii}^{M}(\lambda). \tag{1.113}$$

Now, using the CLT, we know that the last sum converges, for large M, towards $\sigma^2/M \sum_{i=1}^{M} G_{ii}^{M}(\lambda)$. This result is independent of the precise statistics of the H_{0i}, provided their variance is finite.[28] This shows that G_{00} converges for large M towards a well-defined limit G_∞, which obeys the following limit equation:

$$\frac{1}{G_\infty(\lambda)} = \lambda - \sigma^2 G_\infty(\lambda). \tag{1.114}$$

[28] The case of Lévy distributed H_{ij}'s with infinite variance has been investigated in: P. Cizeau, J.-P. Bouchaud, Theory of Lévy matrices, *Physical Review*, **E 50**, 1810 (1994).

The solution to this second-order equation reads:

$$G_\infty(\lambda) = \frac{1}{2\sigma^2}\left[\lambda - \sqrt{\lambda^2 - 4\sigma^2}\right]. \tag{1.115}$$

(The correct solution is chosen to recover the right limit for $\sigma = 0$.) Now, the only way for this quantity to have a non-zero imaginary part when one adds to λ a small imaginary term $i\epsilon$ which tends to zero is that the square root itself is imaginary. The final result for the density of eigenvalues is therefore:

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - \lambda^2} \qquad \text{for } |\lambda| \leq 2\sigma, \tag{1.116}$$

and zero elsewhere. This is the well-known 'semi-circle' law for the density of states, first derived by Wigner. This result can be obtained by a variety of other methods if the distribution of matrix elements is Gaussian. In finance, one often encounters *correlation matrices* $\mathbf{C}$, which have the special property of being positive definite. $\mathbf{C}$ can be written as $\mathbf{C} = \mathbf{H}\mathbf{H}^\dagger$, where $\mathbf{H}^\dagger$ is the matrix transpose of $\mathbf{H}$. In general, $\mathbf{H}$ is a rectangular matrix of size $M \times N$, so $\mathbf{C}$ is $M \times M$. In Chapter 2, M will be the number of assets, and N, the number of observations (days). In the particular case where $N = M$, the eigenvalues of $\mathbf{C}$ are simply obtained from those of $\mathbf{H}$ by squaring them:

$$\lambda_C = \lambda_{\mathrm{H}}^2. \tag{1.117}$$

If one assumes that the elements of $\mathbf{H}$ are random variables, the density of eigenvalues of $\mathbf{C}$ can be obtained from:

$$\rho(\lambda_C)\,\mathrm{d}\lambda_C = 2\rho(\lambda_{\mathrm{H}})\,\mathrm{d}\lambda_{\mathrm{H}} \quad \text{for } \lambda_{\mathrm{H}} > 0, \tag{1.118}$$

where the factor of 2 comes from the two solutions $\lambda_{\mathrm{H}} = \pm\sqrt{\lambda_C}$; this then leads to:

$$\rho(\lambda_C) = \frac{1}{2\pi\sigma^2}\sqrt{\frac{4\sigma^2 - \lambda_C}{\lambda_C}} \qquad \text{for } 0 \leq \lambda_C \leq 4\sigma^2, \tag{1.119}$$

and zero elsewhere. For $N \neq M$, a similar formula exists, which we shall use in the following. In the limit $N, M \to \infty$, with a fixed ratio $Q = N/M \geq 1$, one has:[29]

$$\begin{aligned} \rho(\lambda_C) &= \frac{Q}{2\pi\sigma^2}\frac{\sqrt{(\lambda_{\max} - \lambda_C)(\lambda_C - \lambda_{\min})}}{\lambda_C}, \\ \lambda_{\min}^{\max} &= \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}), \end{aligned} \tag{1.120}$$

with $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ and where σ^2/N is the variance of the elements of $\mathbf{H}$,

[29] A derivation of Eq. (1.120) is given in Appendix B. See also: A. Edelmann, Eigenvalues and condition numbers of random matrices, *SIAM Journal of Matrix Analysis and Applications*, **9**, 543 (1988).

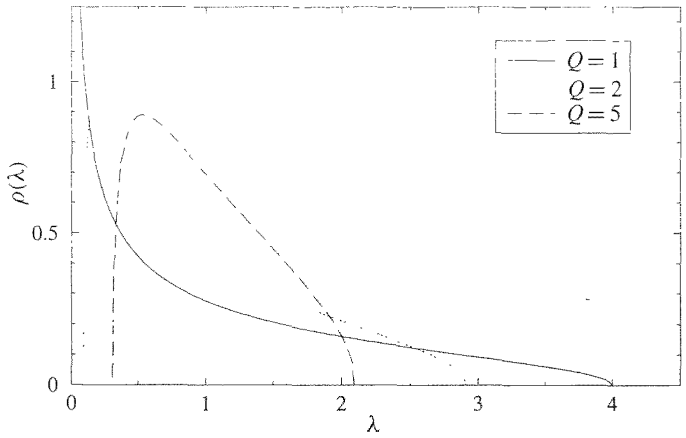

Fig. 1.10. Graph of Eq. (1.120) for $Q = 1, 2$ and 5.

equivalently σ^2 is the average eigenvalue of **C**. This form is actually also valid for $Q < 1$, except that there appears a finite fraction of strictly zero eigenvalues, of weight $1 - Q$ (Fig. 1.10).

The most important features predicted by Eq. (1.120) are:

- The fact that the lower 'edge' of the spectrum is positive (except for $Q = 1$); there is therefore no eigenvalue between 0 and $\lambda_{\min}$. Near this edge, the density of eigenvalues exhibits a sharp maximum, except in the limit $Q = 1$ ($\lambda_{\min} = 0$) where it diverges as $\sim 1/\sqrt{\lambda}$.
- The density of eigenvalues also vanishes above a certain upper edge $\lambda_{\max}$.

Note that all the above results are only valid in the limit $N \to \infty$. For finite N, the singularities present at both edges are smoothed: the edges become somewhat blurred, with a small probability of finding eigenvalues above $\lambda_{\max}$ and below $\lambda_{\min}$, which goes to zero when N becomes large.[30]

In Chapter 2, we will compare the empirical distribution of the eigenvalues of the correlation matrix of stocks corresponding to different markets with the theoretical prediction given by Eq. (1.120).

[30] See e.g. M. J. Bowick, E. Brézin, Universal scaling of the tails of the density of eigenvalues in random matrix models, *Physics Letters*, **B268**, 21 (1991).

1.9 Appendix A: non-stationarity and anomalous kurtosis

In this appendix, we calculate the average kurtosis of the sum $\sum_{i=1}^N \delta x_i$, assuming that the δx_i can be written as $\sigma_i \epsilon_i$. The σ_i's are correlated as:

$$\overline{(D_k - \overline{D})(D_\ell - \overline{D})} = \overline{D}^2 g(|\ell - k|) \qquad D_k \propto \sigma_k^2. \tag{1.121}$$

Let us first compute $\overline{\left\langle \left(\sum_{i=1}^N \delta x_i\right)^4 \right\rangle}$, where $\langle \cdots \rangle$ means an average over the ϵ_i's and the overline means an average over the σ_i's. If $\langle \epsilon_i \rangle = 0$, and $\langle \epsilon_i \epsilon_j \rangle = 0$ for $i \neq j$, one finds:

$$\begin{aligned}\overline{\left\langle \left(\sum_{i,j,k,l=1}^N \delta x_i \delta x_j \delta x_k \delta x_l \right) \right\rangle} &= \sum_{i=1}^N \overline{\langle \delta x_i^4 \rangle} + 3 \sum_{i \neq j=1}^N \overline{\langle \delta x_i^2 \rangle \langle \delta x_j^2 \rangle} \\ &= (3 + \kappa_0) \sum_{i=1}^N \overline{\langle \delta x_i^2 \rangle^2} + 3 \sum_{i \neq j = 1}^N \overline{\langle \delta x_i^2 \rangle \langle \delta x_j^2 \rangle},\end{aligned} \tag{1.122}$$

where we have used the definition of κ_0 (the kurtosis of ϵ). On the other hand, one must estimate $\overline{\left\langle \left(\sum_{i=1}^N \delta x_i\right)^2 \right\rangle}^2$. One finds:

$$\overline{\left\langle \left(\sum_{i=1}^N \delta x_i \right)^2 \right\rangle}^2 = \sum_{i,j=1}^N \overline{\langle \delta x_i^2 \rangle}\; \overline{\langle \delta x_j^2 \rangle}. \tag{1.123}$$

Gathering the different terms and using the definition Eq. (1.121), one finally establishes the following general relation:

$$\kappa_N = \frac{1}{N^2 \overline{D}^2} \left[N\overline{D}^2 (3 + \kappa_0)(1 + g(0)) - 3N\overline{D}^2 + 3\overline{D}^2 \sum_{i \neq j=1}^N g(|i - j|) \right], \tag{1.124}$$

or:

$$\kappa_N = \frac{1}{N} \left[\kappa_0 + (3 + \kappa_0) g(0) + 6 \sum_{\ell=1}^N \left(1 - \frac{\ell}{N}\right) g(\ell) \right]. \tag{1.125}$$

1.10 Appendix B: density of eigenvalues for random correlation matrices

This very technical appendix aims at giving a few of the steps of the computation needed to establish Eq. (1.120). One starts from the following representation of the resolvent $G(\lambda)$:

$$G(\lambda) = \sum_\alpha \frac{1}{\lambda - \lambda_\alpha} = \frac{\partial}{\partial \lambda} \log \prod_\alpha (\lambda - \lambda_\alpha) = \frac{\partial}{\partial \lambda} \log \det(\lambda \mathbf{1} - \mathbf{C}) \equiv \frac{\partial}{\partial \lambda} Z(\lambda). \tag{1.126}$$

Using the following representation for the determinant of a symmetrical matrix $\mathbf{A}$:

$$[\det \mathbf{A}]^{-1/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^M \int \exp\left[-\frac{1}{2}\sum_{i,j=1}^{M} \varphi_i \varphi_j A_{ij}\right] \prod_{i=1}^{M} \mathrm{d}\varphi_i. \tag{1.127}$$

we find, in the case where $\mathbf{C} = \mathbf{H}\mathbf{H}^{\dagger}$:

$$Z(\lambda) = -2\log \int \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{M}\varphi_i^2 - \frac{1}{2}\sum_{i,j=1}^{M}\sum_{k=1}^{N}\varphi_i\varphi_j H_{ik}H_{jk}\right]\prod_{i=1}^{M}\left(\frac{\mathrm{d}\varphi_i}{\sqrt{2\pi}}\right). \tag{1.128}$$

The claim is that the quantity $G(\lambda)$ is self-averaging in the large M limit. So in order to perform the computation we can replace $G(\lambda)$ for a specific realization of $\mathbf{H}$ by the average over an ensemble of $\mathbf{H}$. In fact, one can in the present case show that the average of the logarithm is equal, in the large N limit, to the logarithm of the average. This is not true in general, and one has to use the so-called 'replica trick' to deal with this problem: this amounts to taking the nth power of the quantity to be averaged (corresponding to n copies (replicas) of the system) and to let n go to zero at the end of the computation.[31]

We assume that the $M \times N$ variables H_{ik} are iid Gaussian variable with zero mean and variance σ^2/N. To perform the average of the H_{ik}, we notice that the integration measure generates a term $\exp -M\sum H_{ik}^2/(2\sigma^2)$ that combines with the $H_{ik}H_{jk}$ term above. The summation over the index k doesn't play any role and we get N copies of the result with the index k dropped. The Gaussian integral over H_{io} gives the square-root of the ratio of the determinant of $[M\delta_{ij}/\sigma^2]$ and $[M\delta_{ij}/\sigma^2 - \varphi_i\varphi_j]$:

$$\left\langle \exp\left[-\frac{1}{2}\sum_{i,j=1}^{M}\sum_{k=1}^{N}\varphi_i\varphi_j H_{ik}H_{jk}\right]\right\rangle = \left(1 - \frac{\sigma^2}{N}\sum\varphi_i^2\right)^{-N/2}. \tag{1.129}$$

We then introduce a variable $q \equiv \sigma^2 \sum \varphi_i^2/N$ which we fix using an integral representation of the delta function:

$$\delta\left(q - \sigma^2\sum\varphi_i^2/N\right) = \int \frac{1}{2\pi}\exp\left[\mathrm{i}\zeta\left(q - \sigma^2\sum\varphi_i^2/N\right)\right]\mathrm{d}\zeta. \tag{1.130}$$

After performing the integral over the φ_i's and writing $z = 2\mathrm{i}\zeta/N$, we find:

$$\begin{aligned} Z(\lambda) &= -2\log\frac{N}{4\pi}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}\int_{-\infty}^{\infty}\exp\left[-\frac{M}{2}(\log(\lambda - \sigma^2 z) + \right. \\ &\qquad \left. Q\log(1-q) + Qqz)\right]\mathrm{d}q\,\mathrm{d}z, \end{aligned} \tag{1.131}$$

[31] For more details on this technique, see, for example, M. Mézard, G. Parisi, M. A. Virasoro, *Spin Glasses and Beyond*, World Scientific, Singapore, 1987.

where $Q = N/M$. The integrals over z and q are performed by the saddle point method, leading to the following equations:

$$Qq = \frac{\sigma^2}{\lambda - \sigma^2 z} \quad \text{and} \quad z = \frac{1}{1-q}. \tag{1.132}$$

The solution in terms of $q(\lambda)$ is:

$$q(\lambda) = \frac{\sigma^2(1-Q) + Q\lambda \pm \sqrt{(\sigma^2(1-Q)+Q\lambda)^2 - 4\sigma^2 Q\lambda}}{2Q\lambda}. \tag{1.133}$$

We find $G(\lambda)$ by differentiating Eq. (1.131) with respect to λ. The computation is greatly simplified if we notice that at the saddle point the partial derivatives with respect to the functions $q(\lambda)$ and $z(\lambda)$ are zero by construction. One finally finds:

$$G(\lambda) = \frac{M}{\lambda - \sigma^2 z(\lambda)} = \frac{MQq(\lambda)}{\sigma^2}. \tag{1.134}$$

We can now use Eq. (1.110) and take the imaginary part of $G(\lambda)$ to find the density of eigenvalues:

$$\rho(\lambda) = \frac{\sqrt{4\sigma^2 Q\lambda - (\sigma^2(1-Q)+Q\lambda)^2}}{2\pi\lambda\sigma^2}, \tag{1.135}$$

which is identical to Eq. (1.120).

1.11 References

Introduction to probabilities

W. Feller, *An Introduction to Probability Theory and its Applications*, Wiley, New York, 1971.

Extreme value statistics

E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.

Sum of random variables, Lévy distributions

B. V. Gnedenko, A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison Wesley, Cambridge, MA, 1954.
P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier Villars, Paris, 1937–1954.
G. Samorodnitsky, M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman & Hall, New York, 1994.

Broad distributions in natural sciences and in finance

B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
B. B. Mandelbrot, *Fractals and Scaling in Finance*, Springer, New York, 1997.

Broad distributions in physics

E. Montroll, M. Shlesinger, The Wonderful World of Random Walks, in *Nonequilibrium Phenomena II, From Stochastics to Hydrodynamics*, Studies in Statistical Mechanics XI (J. L. Lebowitz, E. W. Montroll, eds) North-Holland, Amsterdam, 1984.
J.-P. Bouchaud, A. Georges, Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications, *Physics Reports*, **195**, 127 (1990).
G. Zaslavsky, U. Frisch, M. Shlesinger, *Lévy Flights and Related Topics in Physics*, Lecture Notes in Physics **450**, Springer, New York, 1995.

Fractional Brownian motion

B. B. Mandelbrot, J. W. Van Ness, Fractional Brownian Motion, fractional noises and applications, *SIAM Review*, **10**, 422 (1968).

Theory of random matrices

M. L. Mehta, *Random Matrix Theory*, Academic Press, New York, 1995.
O. Bohigas, M. J. Giannoni, *Mathematical and Computational Methods in Nuclear Physics*, Lecture Notes in Physics, vol. 209, Springer-Verlag, New York, 1983.

2

Statistics of real prices

Le marché, à son insu, obéit à une loi qui le domine : la loi de la probabilité.[1]

(Bachelier, *Théorie de la spéculation.*)

2.1 Aim of the chapter

The easy access to enormous financial databases, containing thousands of asset time series, sampled at a frequency of minutes or sometimes seconds, allows one to investigate in detail the statistical features of the time evolution of financial assets. The description of any kind of data, be it of physical, biological, or financial origin, requires however an interpretation framework, needed to order and give a meaning to the observations. To describe necessarily means to simplify, and even sometimes betray: the aim of any empirical science is to approach reality progressively, through successive improved approximations.

The goal of the present chapter is to present in this spirit the statistical properties of financial time series. We shall propose some plausible mathematical modelling, as faithful as possible (though imperfect) of the observed properties of these time series. The models we discuss are however not the only possible models; the available data is often not sufficiently accurate to distinguish, say, between a truncated Lévy distribution and a Student distribution. The choice between the two is then guided by mathematical convenience. In this respect, it is interesting to note that the word 'modelling' has two rather different meanings within the scientific community. The first one, often used in applied mathematics, engineering sciences and financial mathematics, means that one represents reality using appropriate mathematical formulae. This is the scope of the present chapter. The second, more common in the physical sciences, is perhaps more ambitious: it aims at finding a set of plausible causes sufficient to explain the observed phenomena,

[1] The market, without knowing it, obeys a law which overwhelms it: the law of probability.

and therefore, ultimately, to justify the chosen mathematical description. We will however only discuss in a cursory way the 'microscopic' mechanisms of price formation and evolution, of adaptive traders' strategies, herding behaviour between traders, feedback of price variations onto themselves, etc., which are certainly at the origin of the interesting statistics that we shall report below. We feel that this aspect of the problem is still in its infancy, and will evolve rapidly in the coming years. We briefly mention, at the end of this chapter, two simple models of herding and feedback, and give references of several very recent articles.

We shall describe several types of market:

- Very liquid, 'mature' markets of which we take three examples: a US stock index (S&P 500), an exchange rate (DEM/$), and a long-term interest rate index (the German Bund);
- Very volatile markets, such as emerging markets like the Mexican peso;
- Volatility markets: through option markets, the volatility of an asset (which is empirically found to be time dependent) can be seen as a price which is quoted on markets (see Chapter 4);
- Interest rate markets, which give fluctuating prices to loans of different maturities, between which special types of correlations must however exist.

We chose to limit our study to fluctuations taking place on rather short time scales (typically from minutes to months). For longer time scales, the available data-set is in general too small to be meaningful. From a fundamental point of view, the influence of the average return is negligible for short time scales, but becomes crucial on long time scales. Typically, a stock varies by several per cent within a day, but its average return is, say, 10% per year, or 0.04% per day. Now, the 'average return' of a financial asset appears to be unstable in time: the past return of a stock is seldom a good indicator of future returns. Financial time series are intrinsically non-stationary: new financial products appear and influence the markets, trading techniques evolve with time, as does the number of participants and their access to the markets, etc. This means that taking very long historical data-set to describe the long-term statistics of markets is *a priori* not justified. We will thus avoid this difficult (albeit important) subject of long time scales.

The simplified model that we will present in this chapter, and that will be the starting point of the theory of portfolios and options discussed in later chapters, can be summarized as follows. The variation of price of the asset X between time $t = 0$ and $t = T$ can be decomposed as:

$$x(T) = x_0 + \sum_{k=0}^{N-1} \delta x_k \qquad N = \frac{T}{\tau}, \tag{2.1}$$

where,

- *In a first approximation*, and for T not too large, the price increments δx_k are random variables which are (i) independent as soon as τ is larger than a few tens of minutes (on liquid markets) and (ii) identically distributed, according to a TLD, Eq. (1.23), $P_1(\delta x) = L_\mu^{(t)}(\delta x)$ with a parameter μ approximately equal to $\frac{3}{2}$, *for all markets*.[2] The exponential cut-off appears 'earlier' in the tail for liquid markets, and can be completely absent in less mature markets.

  The results of Chapter 1 concerning sums of random variables, and the convergence towards the Gaussian distribution, allows one to understand the observed 'narrowing' of the tails as the time interval T increases.
- A refined analysis however reveals important systematic deviations from this simple model. In particular, the kurtosis of the distribution of $x(T)-x_0$ decreases more slowly than $1/N$, as it should if the increments δx_k were iid random variables. This suggests a certain form of temporal dependence, of the type discussed in Section 1.7.2. The volatility (or the variance) of the price increments δx is actually itself time dependent: this is the so-called 'heteroskedasticity' phenomenon. As we shall see below, periods of high volatility tend to persist over time, thereby creating long-range higher-order correlations in the price increments. On long time scales, one also observes a systematic dependence of the variance of the price increments on the price x itself. In the case where the RMS of the variables δx grows linearly with x, the model becomes multiplicative, in the sense that one can write:

$$x(T) = x_0 \prod_{k=0}^{N-1} (1 + \eta_k) \qquad N = \frac{T}{\tau}, \tag{2.2}$$

  where the returns η_k have a fixed variance. This model is actually more commonly used in the financial literature. We will show that reality must be described by an intermediate model, which interpolates between a purely additive model, Eq. (2.1), and a multiplicative model, Eq. (2.2).

Studied assets

The chosen stock index is the futures contract on the Standard and Poor's 500 (S&P 500) US stock index, traded on the Chicago Mercantile Exchange (CME). During the time period chosen (from November 1991 to February 1995), the index rose from 375 to 480 points (Fig. 2.1 (top)). Qualitatively, all the conclusions reached on this period of time are more generally valid, although the value of some parameters (such as the volatility) can change significantly from one period to the next.

The exchange rate is the US dollar ($) against the German mark (DEM), which is the most active exchange rate market in the world. During the analysed period, the mark varied

[2] Alternatively, a description in terms of Student distributions is often found to be of comparable quality, with a tail exponent $\mu \sim 3$–5 for the S&P 500, for example.

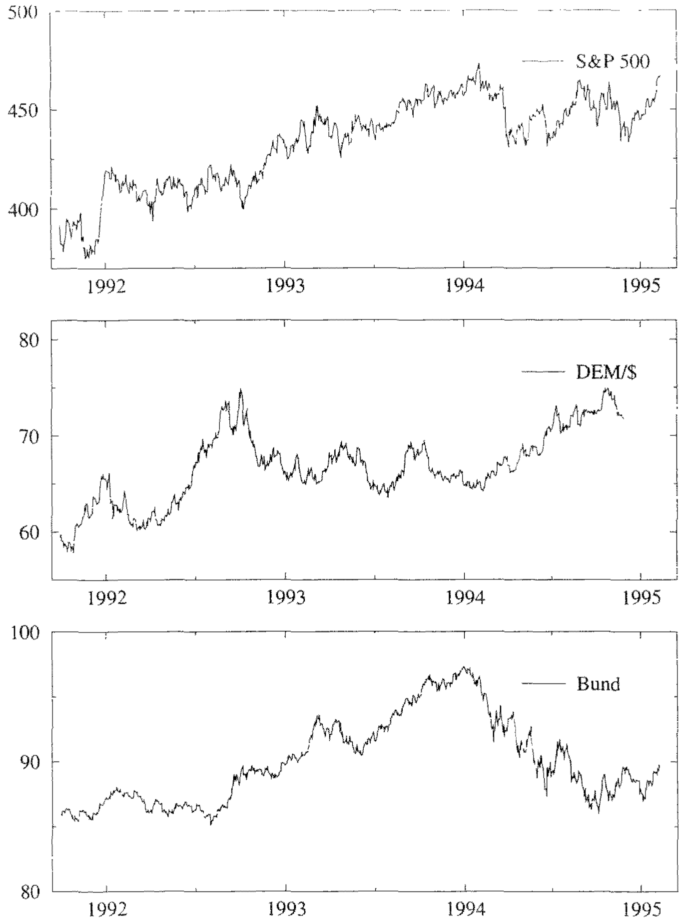

Fig. 2.1. Charts of the studied assets between November 1991 and February 1995. The top chart is the S&P 500, the middle one is the DEM/$, and the bottom one is the long-term German interest rate (Bund).

between 58 and 75 cents (Fig. 2.1 (middle)). Since the interbank settlement prices are not available, we have defined the price as the average between the bid and the ask prices.[3]

Finally, the chosen interest rate index is the futures contract on long-term German bonds (Bund), quoted on the London International Financial Futures and Options Exchange (LIFFE). It is typically varying between 85 and 100 points (Fig. 2.1 (bottom)).

[3] There is, on all financial markets, a difference between the bid price and the ask price for a certain asset at a given instant of time. The difference between the two is called the 'bid/ask spread'. The more liquid a market, the smaller the average spread.

The indices S&P 500 and Bund that we have studied are thus actually futures contracts (cf. Section 4.2). The fluctuations of futures prices follow in general those of the underlying contract and it is reasonable to identify the statistical properties of these two objects. Futures contracts exist with several fixed maturity dates. We have always chosen the most liquid maturity and suppressed the artificial difference of prices when one changes from one maturity to the next (roll). We have also neglected the weak dependence of the futures contracts on the short time interest rate (see Section 4.2): this trend is completely masked by the fluctuations of the underlying contract itself.

2.2 Second-order statistics

2.2.1 Variance, volatility and the additive–multiplicative crossover

In all that follows, the notation δx represents the difference of value of the asset X between two instants separated by a time interval τ:

$$\delta x_k = x(t+\tau) - x(t) \qquad t \equiv k\tau. \tag{2.3}$$

In the whole modern financial literature, it is postulated that the relevant variable is not the increment δx itself, but rather the *return* $\eta = \delta x / x$. It is therefore interesting to study empirically the variance of δx, conditioned to a certain value of the price x itself, which we shall denote $\langle \delta x^2 \rangle|_x$. If the return η is the natural random variable, one should observe that $\sqrt{\langle \delta x^2 \rangle|_x} = \sigma_1 x$, where σ_1 is constant (and equal to the RMS of η). Now, in many instances (Figs 2.2 and 2.4), one rather finds that $\sqrt{\langle \delta x^2 \rangle|_x}$ is independent of x, apart from the case of exchange rates between comparable currencies. The case of the CAC 40 is particularly interesting, since during the period 1991–95, the index went from 1400 to 2100, leaving the absolute volatility nearly constant (if anything, it is seen to decrease with x!).

On longer time scales, however, or when the price x rises substantially, the RMS of δx increases significantly, as to become proportional to x (Fig. 2.3). A way to model this crossover from an additive to a multiplicative behaviour is to postulate that the RMS of the increments progressively (over a time scale T_σ) adapt to the changes of price of x. Schematically, for $T < T_\sigma$, the prices behave additively, whereas for $T > T_\sigma$, multiplicative effects start playing a significant role:[4]

$$\begin{aligned} \langle (x(T) - x_0)^2 \rangle &= DT \quad (T \ll T_\sigma); \\ \left\langle \log^2 \left(\frac{x(T)}{x_0} \right) \right\rangle &= \sigma^2 T \quad (T \gg T_\sigma). \end{aligned} \tag{2.4}$$

[4] In the additive regime, where the variance of the increments can be taken as a constant, we shall write $\langle \delta x^2 \rangle = \sigma_1^2 x_0^2 = D\tau$.

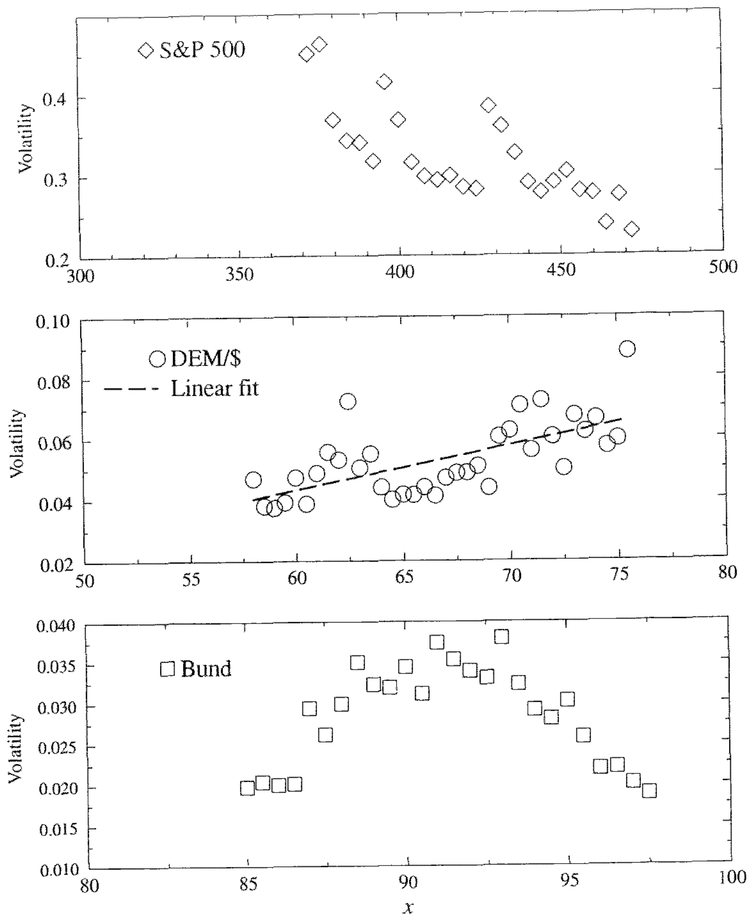

Fig. 2.2. RMS of the increments δx, conditioned to a certain value of the price x, as a function of x, for the three chosen assets. For the chosen period, only the exchange rate DEM/$ conforms to the idea of a multiplicative model: the straight line corresponds to the best fit $\langle \delta x^2 \rangle|_x^{1/2} = \sigma_1 x$. The adequacy of the multiplicative model in this case is related to the symmetry $/DEM $\to$ DEM/$.

On liquid markets, this time scale is on the order of months. A convenient way to model this crossover is to introduce an *additive* random variable $\xi(T)$, and to represent the price $x(T)$ as $x(T) = x_0(1 + \xi(T)/q(T))^{q(T)}$. For $T \ll T_\sigma$, $q \to 1$, the price process is additive, whereas for $T \gg T_\sigma$, $q \to \infty$, which corresponds to the multiplicative limit.

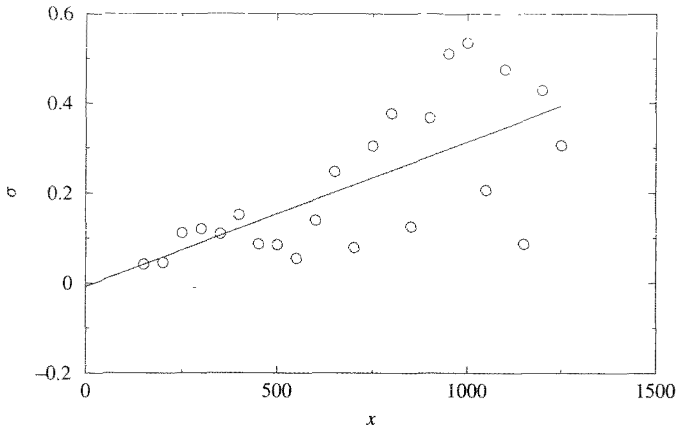

Fig. 2.3. RMS of the increments δx, conditioned to a certain value of the price x, as a function of x, for the S&P 500 for the 1985–98 time period.

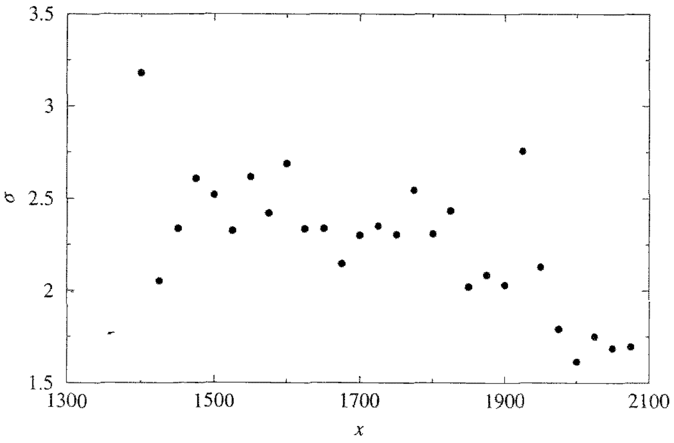

Fig. 2.4. RMS of the increments δx, conditioned to a certain value of the price x, as a function of x, for the CAC 40 index for the 1991–95 period; it is quite clear that during that time period $\langle \delta x^2 \rangle|_x$ was almost independent of x.

2.2.2 Autocorrelation and power spectrum

The simplest quantity, commonly used to measure the correlations between price increments, is the temporal two-point correlation function $C^\tau_{k\ell}$, defined as:[5]

$$C^\tau_{k\ell} = \frac{1}{D\tau}\langle \delta x_k \delta x_\ell \rangle; \qquad D\tau = \langle \delta x_k^2 \rangle. \tag{2.5}$$

[5] In principle, one should subtract the average value $\langle \delta x \rangle = m\tau = m_1$ from δx. However, if τ is small (for example equal to a day), $m\tau$ is completely negligible compared to $\sqrt{D\tau}$.

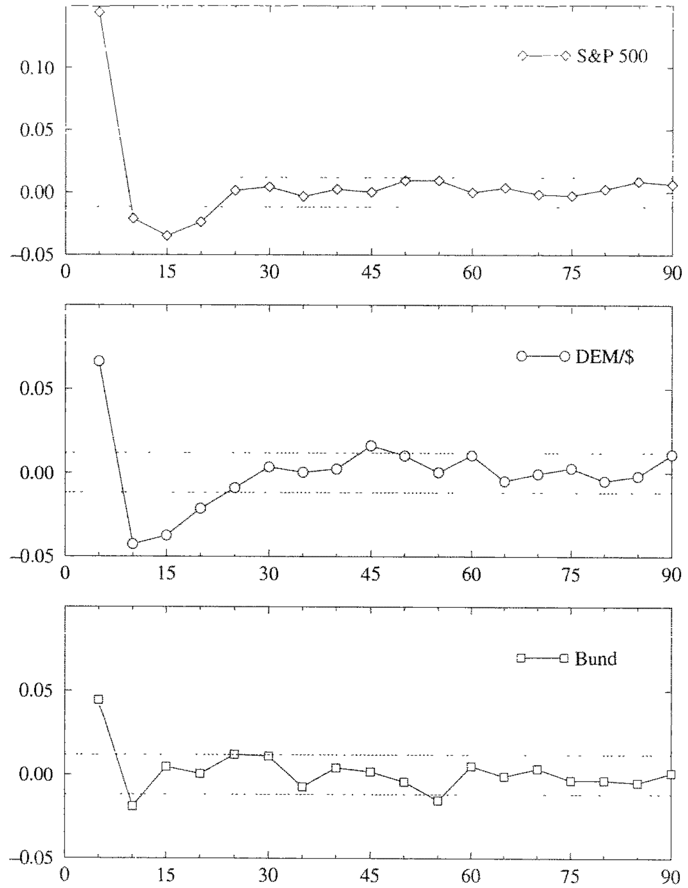

Fig. 2.5. Normalized correlation function $C^{\tau}_{k\ell}$ for the three chosen assets, as a function of the time difference $|k - l|\tau$, and for $\tau = 5$ min. Up to 30 min, some weak but significant correlations do exist (of amplitude ~ 0.05). Beyond 30 min, however, the two-point correlations are not statistically significant.

Figure 2.5 shows this correlation function for the three chosen assets, and for $\tau = 5$ min. For uncorrelated increments, the correlation function $C^{\tau}_{k\ell}$ should be equal to zero for $k \neq l$, with an RMS equal to $\sigma = 1/\sqrt{N}$, where N is the number of independent points used in the computation. Figure 2.5 also shows the 3σ error bars. We conclude that beyond 30 min, the two-point correlation function cannot be distinguished from zero. On less liquid markets, however, this correlation time is longer. On the US stock market, for example, this correlation time has significantly decreased between the 1960s and the 1990s.

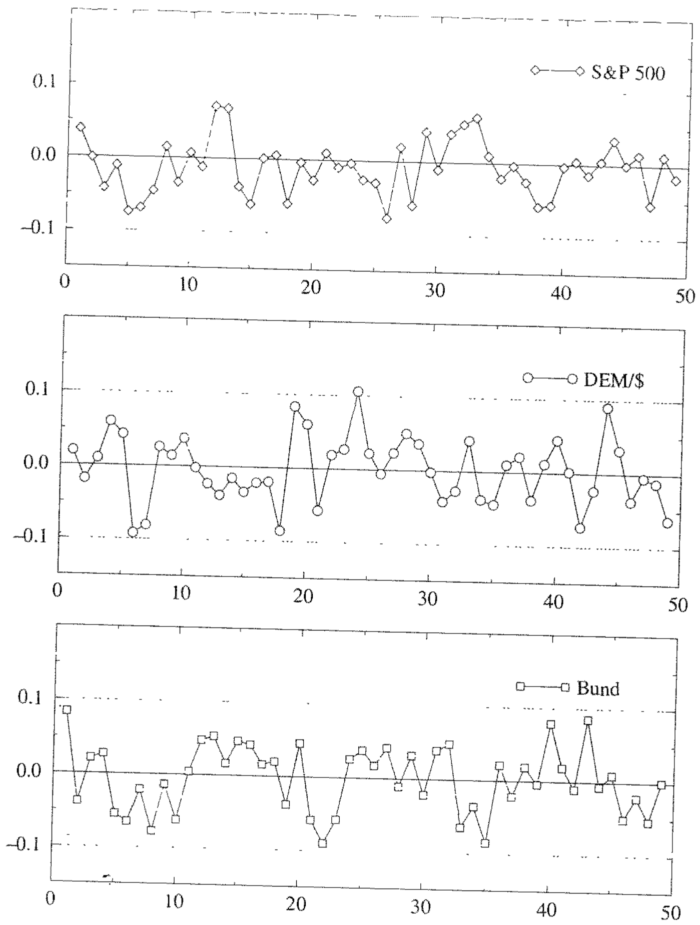

Fig. 2.6. Normalized correlation function $C^{\tau}_{k\ell}$ for the three chosen assets, as a function of the time difference $|k - l|\tau$, now on a daily basis, $\tau = 1$ day. The two horizontal lines at ± 0.1 correspond to a 3σ error bar. No significant correlations can be measured.

On very short time scales, however, weak but significant correlations do exist. These correlations are however too small to allow profit making: the potential return is smaller than the transaction costs involved for such a high-frequency trading strategy, even for the operators having direct access to the markets (cf. Section 4.1.2). Conversely, if the transaction costs are high, one may expect significant correlations to exist on longer time scales.

We have performed the same analysis for the daily increments of the three chosen assets ($\tau = 1$ day). Figure 2.6 reveals that the correlation function is

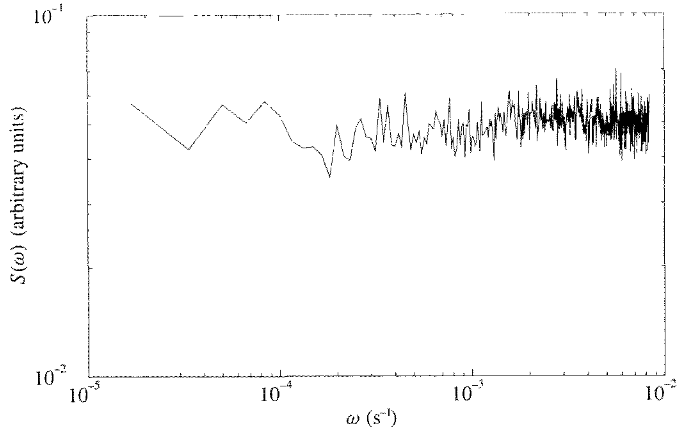

Fig. 2.7. Power spectrum $\mathcal{S}(\omega)$ of the time series DEM/\$, as a function of the frequency ω. The spectrum is flat: for this reason one often speaks of white noise, where all the frequencies are represented with equal weights. This corresponds to uncorrelated increments.

always within 3σ of zero, confirming that the daily increments are not significantly correlated.

Power spectrum

Let us briefly mention another equivalent way of presenting the same results, using the so-called power spectrum, defined as:

$$\mathcal{S}(\omega) = \frac{1}{N}\left\langle \sum_{k,\ell=1}^{N} \delta x_k \delta x_\ell e^{i\omega(k-l)} \right\rangle. \tag{2.6}$$

The case of uncorrelated increments leads to a flat power spectrum, $\mathcal{S}(\omega) = \mathcal{S}_0$. Figure 2.7 shows the power spectrum of the DEM/\$ time series, where no significant structure appears.

2.3 Temporal evolution of fluctuations

2.3.1 Temporal evolution of probability distributions

The results of the previous section are compatible with the simplest scenario where the price increments δx_k are, beyond a certain correlation time, independent random variables. A much finer test of this assumption consists in studying directly the probability distributions of the price increments $x_N - x_0 = \sum_{k=0}^{N-1} \delta x_k$ on different time scales $N = T/\tau$. If the increments are independent, then the distributions on different time scales can be obtained from the one pertaining to

Table 2.1. Value of the parameters A and α^{-1}, as obtained by fitting the data with a symmetric TLD $L_\mu^{(t)}$, of index $\mu = \frac{3}{2}$. Note that both A and α^{-1} have the dimension of a price variation δx_1, and therefore directly characterize the nature of the statistical fluctuations. The other columns compare the RMS and the kurtosis of the fluctuations, as directly measured on the data, or via the formulae, Eqs (1.94), (1.95). Note that in the case DEM/\$, the studied variable is $100\delta x/x$. In this last case, the fit with $\mu = 1.5$ is not very good: the calculated kurtosis is found to be too high. A better fit is obtained with $\mu = 1.2$

Asset			Variance σ_1^2		Kurtosis κ_1	
	A	α^{-1}	Measured	Computed	Measured	Computed
S&P 500	0.22	2.21	0.280	0.279	12.7	13.1
Bund	0.0091	0.275	0.002 40	0.002 42	20.4	23.5
DEM/\$	0.0447	0.96	0.0163	0.0164	20.5	41.9

the elementary time scale τ (chosen to be larger than the correlation time). More precisely (see Section 1.5.1), one should have $P(x, N) = [P_1(\delta x_1)]^{*N}$.

The elementary distribution P_1

The elementary cumulative probability distribution $\mathcal{P}_{1>}(\delta x)$ is represented in Figures 2.8, 2.9 and 2.10. One should notice that the tail of the distribution is broad, in any case much broader than a Gaussian. A fit using a truncated Lévy distribution of index $\mu = \frac{3}{2}$, as given by Eq. (1.23), is quite satisfying.[6] The corresponding parameters A and α are given in Table 2.1 (For $\mu = \frac{3}{2}$, the relation between A and $a_{3/2}$ reads: $a_{3/2} = 2\sqrt{2\pi}A^{3/2}/3$.) Alternatively, as shown in Figure 1.5, a fit using a Student distribution would also be acceptable.

We have chosen to fix the value of μ to $\frac{3}{2}$. This reduces the number of adjustable parameters, and is guided by the following observations:

- A large number of empirical studies on the use of Lévy distributions to fit the financial market fluctuations report values of μ in the range 1.6–1.8. However, in the absence of truncation (i.e. with $\alpha = 0$), the fit overestimates the tails of the distribution. Choosing a higher value of μ partly corrects for this effect, since it leads to a thinner tail.
- If the exponent μ is left as a free parameter, it is in many cases found to be in the range 1.4–1.6, although sometimes smaller, as in the case of the DEM/\$ ($\mu \simeq 1.2$).

[6] A more refined study of the tails actually reveals the existence of a small asymmetry, which we neglect here. Therefore, the skewness λ_3 is taken to be zero.

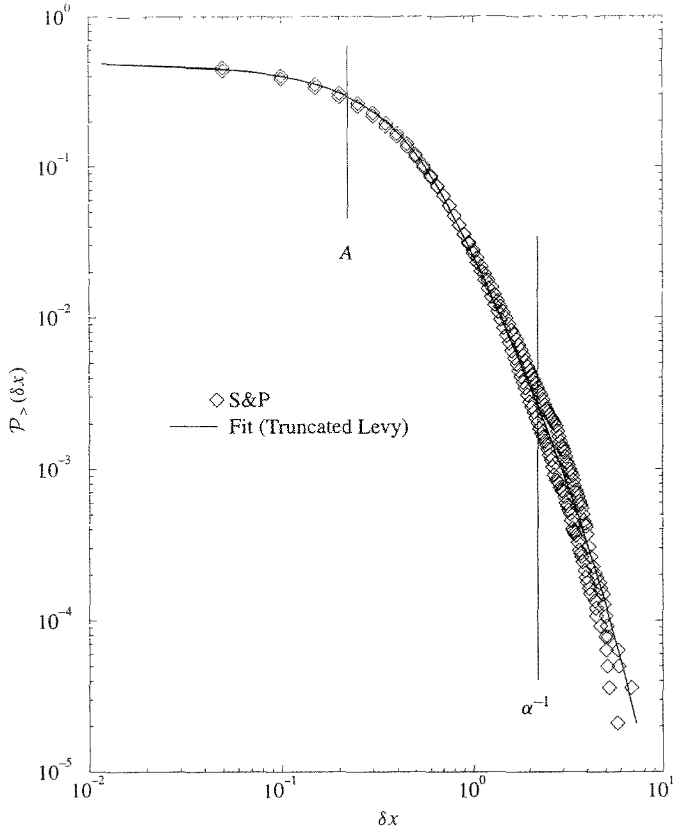

Fig. 2.8. Elementary cumulative distribution $\mathcal{P}_{1>}(\delta x)$ (for $\delta x > 0$) and $\mathcal{P}_{1<}(\delta x)$ (for $\delta x < 0$), for the S&P 500, with $\tau = 15$ min. The thick line corresponds to the best fit using a symmetric TLD $L_\mu^{(t)}$, of index $\mu = \frac{3}{2}$. We have also shown on the same graph the values of the parameters A and α^{-1} as obtained by the fit.

- The particular value $\mu = \frac{3}{2}$ has a simple theoretical interpretation, which we shall briefly present in Section 2.8.

In order to characterize a probability distribution using empirical data, it is always better to work with the cumulative distribution function rather than with the distribution density. To obtain the latter, one indeed has to choose a certain width for the bins in order to construct frequency histograms, or to smooth the data using, for example, a Gaussian with a certain width. Even when this width is carefully chosen, part of the information is lost. It is furthermore difficult to characterize the tails of the distribution, corresponding to rare events, since most bins in this region are empty. On the other hand, the construction of the cumulative distribution does not require to choose a bin width. The trick is to order the observed data according to their rank, for example in decreasing order. The value x_k of the

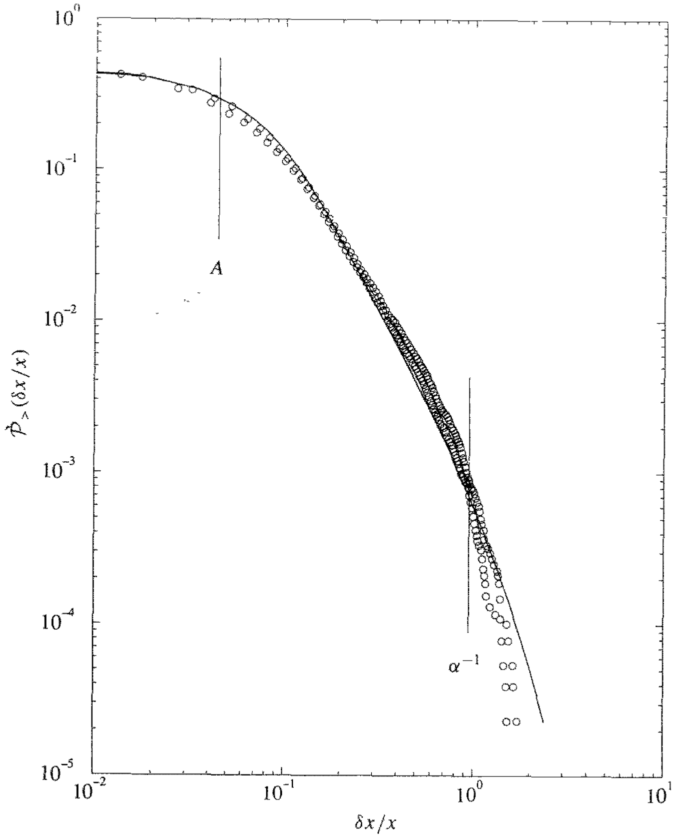

Fig. 2.9. Elementary cumulative distribution for the DEM/\$, for $\tau = 15$ min, and best fit using a symmetric TLD $L_\mu^{(t)}$, of index $\mu = \frac{3}{2}$. In this case, it is rather $100\delta x/x$ that has been considered. The fit is not very good, and would have been better with a smaller value of $\mu \sim 1.2$. This increases the weight of very small variations.

kth variable (out of N) is then such that:

$$\mathcal{P}_>(x_k) = \frac{k}{N+1}. \tag{2.7}$$

This result comes from the following observation: if one draws an $(N+1)$th random variable from the same distribution, there is an a priori equal probability $1/N+1$ that it falls within any of the $N+1$ intervals defined by the previously drawn variables. The probability that it falls above the kth one, x_k is therefore equal to the number of intervals beyond x_k, which is equal to k, times $1/N+1$. This is also equal, by definition, to $\mathcal{P}_>(x_k)$. (See also the discussion in Section 1.4, and Eq. (1.45)). Since the rare events part of the distribution is a particular interest, it is convenient to choose a logarithmic scale for the probabilities. Furthermore, in order to check visually the symmetry of the probability

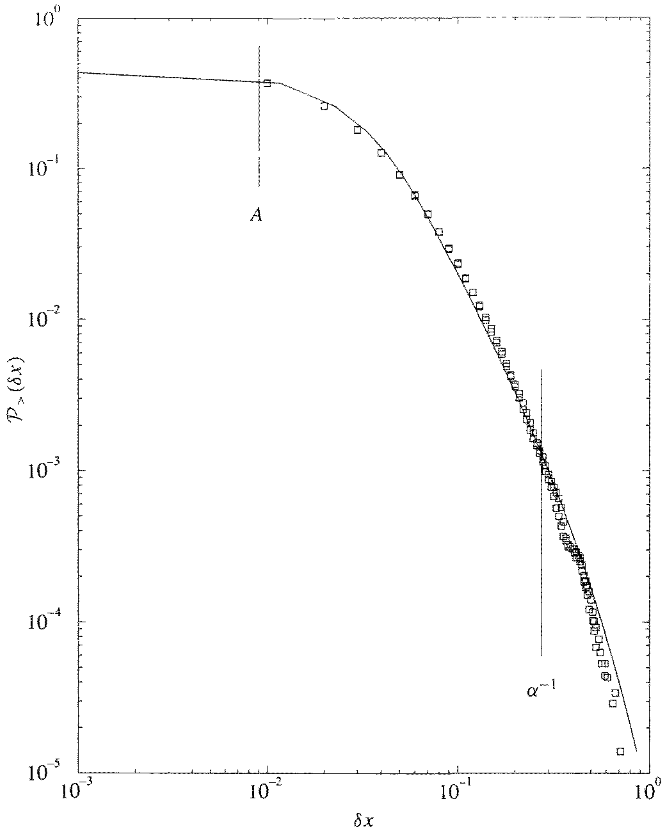

Fig. 2.10. Elementary cumulative distribution for the Bund, for $\tau = 15$ min, and best fit using a symmetric TLD $L_\mu^{(t)}$, of index $\mu = \frac{3}{2}$.

distributions, we have systematically used $\mathcal{P}_<(-\delta x)$ for the negative increments, and $\mathcal{P}_>(\delta x)$ for positive δx.

Maximum likelihood

Suppose that one observes a series of N realizations of the random iid variable X, $\{x_1, x_2, \ldots, x_N\}$, drawn with an unknown distribution that one would like to parameterize, for simplicity, by a single parameter μ. If $P_\mu(x)$ denotes the corresponding probability distribution, the a priori probability to observe the particular series $\{x_1, x_2, \ldots, x_N\}$ is proportional to:

$$P_\mu(x_1)P_\mu(x_2)\ldots P_\mu(x_N). \tag{2.8}$$

The most likely value μ^* of μ is such that this a priori probability is maximized. Taking for example $P_\mu(x)$ to be a power-law distribution:

$$P_\mu(x) = \frac{\mu x_0^\mu}{x^{1+\mu}} \qquad x > x_0, \tag{2.9}$$

(with x_0 known), one has:

$$P_\mu(x_1)P_\mu(x_2)\ldots P_\mu(x_N) \propto \mathrm{e}^{N\log\mu + N\mu\log x_0 - (1+\mu)\sum_{i=1}^N \log x_i}. \tag{2.10}$$

The equation fixing μ^* is thus, in this case:

$$\frac{N}{\mu^*} + N\log x_0 - \sum_{i=1}^{N}\log x_i = 0 \Rightarrow \mu^* = \frac{N}{\sum_{i=1}^N \log(x_i/x_0)}. \tag{2.11}$$

This method can be generalized to several parameters. In the above example, if x_0 is unknown, its most likely value is simply given by: $x_0 = \min\{x_1, x_2, \ldots, x_N\}$.

Convolutions

The parameterization of $P_1(\delta x)$ as a TLD allows one to reconstruct the distribution of price increments for all time intervals $T = N\tau$, if one assumes that the increments are iid random variables. As discussed in Chapter 1, one then has $P(\delta x, N) = [P_1(\delta x_1)]^{\star N}$. Figure 2.11 shows the cumulative distribution for $T = 1$ hour, 1 day and 5 days, reconstructed from the one at 15 min, according to the simple iid hypothesis. The symbols show empirical data corresponding to the same time intervals. The agreement is acceptable; one notices in particular the progressive deformation of $P(\delta x, N)$ towards a Gaussian for large N. The evolution of the variance and of the kurtosis as a function of N is given in Table 2.2, and compared with the results that one would observe if the simple convolution rule was obeyed, i.e. $\sigma_N^2 = N\sigma_1^2$ and $\kappa_N = \kappa_1/N$. For these liquid assets, the time scale $T^* = \kappa_1\tau$ which sets the convergence towards the Gaussian is on the order of days. However, it is clear from Table 2.2 that this convergence is slower than it ought to be: κ_N decreases much more slowly than the $1/N$ behaviour predicted by an iid hypothesis. A closer look at Figure 2.11 also reveals systematic deviations: for example the tails at 5 days are distinctively fatter than they should be.

Tails, what tails?

The asymptotic tails of the distributions $P(\delta x, N)$ are approximately exponential for all N. This is particularly clear for $T = N\tau = 1$ day, as illustrated in Figure 2.12 in a semi-logarithmic plot. However, as mentioned in Section 1.3.4 and in the above paragraph, the distribution of price changes can also be satisfactorily fitted using Student distributions (which have power-law tails) with rather high exponents. In some cases, for example the distribution of losses of the S&P 500 (Fig. 2.12), one sees a slight upward bend in the plot of $\mathcal{P}_>(x)$ versus x

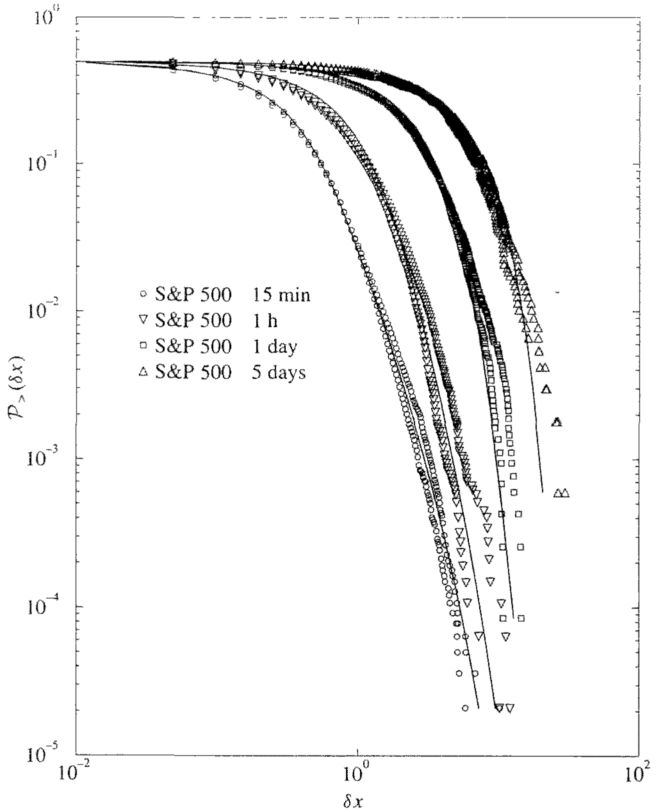

Fig. 2.11. Evolution of the distribution of the price increments of the S&P 500, $P(\delta x, N)$ (symbols), compared with the result obtained by a simple convolution of the elementary distribution $P_1(\delta x_1)$ (dark lines). The width and the general shape of the curves are rather well reproduced within this simple convolution rule. However, systematic deviations can be observed, in particular for large $|\delta x|$. This is also reflected by the fact that the kurtosis κ_N decreases more slowly than κ_1/N, cf. Table 2.2.

in a linear–log plot. This indeed suggests that the decay could be slower than exponential. Many authors have proposed that the tails of the distribution of price changes is a *stretched exponential* $\exp(-|\delta x|^c)$ with $c < 1$,[7] or even a power-law with an exponent μ in the range 3–5.[8] For example, the most likely value of μ

[7] See: J. Laherrère, D. Sornette, Stretched exponential distributions in nature and in economy, *European Journal of Physics*, **B 2**, 525 (1998).

[8] See e.g. M. M. Dacorogna, U. A. Muller, O. V. Pictet, C. G. de Vries, The distribution of extremal exchange rate returns in extremely large data sets, Olsen and Associate working paper (1995), available at http://www.olsen.ch; F. Longin, The asymptotic distribution of extreme stock market returns, *Journal of Business* **69**, 383 (1996); P. Gopikrishnan, M. Meyer, L. A. Amaral, H. E. Stanley, Inverse cubic law for the distribution of stock price variations, *European Journal of Physics*, **B 3**, 139 (1998).

Table 2.2. Variance and kurtosis of the distributions $P(\delta x, N)$ measured or computed from the variance and kurtosis at time scale τ by assuming a simple convolution rule, leading to $\sigma_N^2 = N\sigma_1^2$ and $\kappa_N = \kappa_1/N$. The kurtosis at scale N is systematically too large, cf. Section 2.4. We have used $N = 4$ for $T = 1$ hour, $N = 28$ for $T = 1$ day and $N = 140$ for $T = 5$ days

Asset	Variance σ_N^2 Measured	Variance σ_N^2 Computed	Kurtosis κ_N Measured	Kurtosis κ_N Computed
S&P 500 ($T = 1$ h)	1.06	1.12	6.65	3.18
Bund ($T = 1$ h)	9.49×10^{-3}	9.68×10^{-3}	10.9	5.88
DEM/$ ($T = 1$ h)	6.03×10^{-2}	6.56×10^{-2}	7.20	5.11
S&P 500 ($T = 1$ day)	7.97	7.84	1.79	0.45
Bund ($T = 1$ day)	6.80×10^{-2}	6.76×10^{-2}	4.24	0.84
DEM/$ ($T = 1$ day)	0.477	0.459	1.68	0.73
S&P 500 ($T = 5$ days)	38.6	39.20	1.85	0.09
Bund ($T = 5$ days)	0.341	0.338	1.72	0.17
DEM/$ ($T = 5$ days)	2.52	2.30	0.91	0.15

using a Student distribution to fit the daily variations of the S&P in the period 1991–95 is $\mu = 5$. Even if it is rather hard to distinguish empirically between an exponential and a high power-law, this question is very important theoretically. In particular, the existence of a finite kurtosis requires μ to be larger than 4. As far as applications to risk control, for example, are concerned, the difference between the extrapolated values of the risk using an exponential or a high power-law fit of the tails of the distribution is significant, but not dramatic. For example, fitting the tail of an exponential distribution by a power-law, using 1000 days, leads to an effective exponent $\mu \simeq 4$. An extrapolation to the most probable drop in 10 000 days overestimates the true figure by a factor 1.3. In any case, the amplitude of very large crashes observed in the century are beyond any reasonable extrapolation of the tails, whether one uses an exponential or a high power-law. The *a priori* probability of observing a 22% drop in a single day, as happened on the New York Stock Exchange in October 1987, is found in any case to be much smaller than 10^{-4} per day, that is, once every 40 years. This suggests that major crashes are governed by a specific amplification mechanism, which drives these events outside the scope of a purely statistical analysis, and require a specific theoretical description.[9]

[9] On this point, see A. Johansen, D. Sornette, Stock market crashes are outliers, *European Journal of Physics*, **B 1**, 141 (1998), and J.-P. Bouchaud, R. Cont, *European Journal of Physics*, **B 6**, 543 (1998).

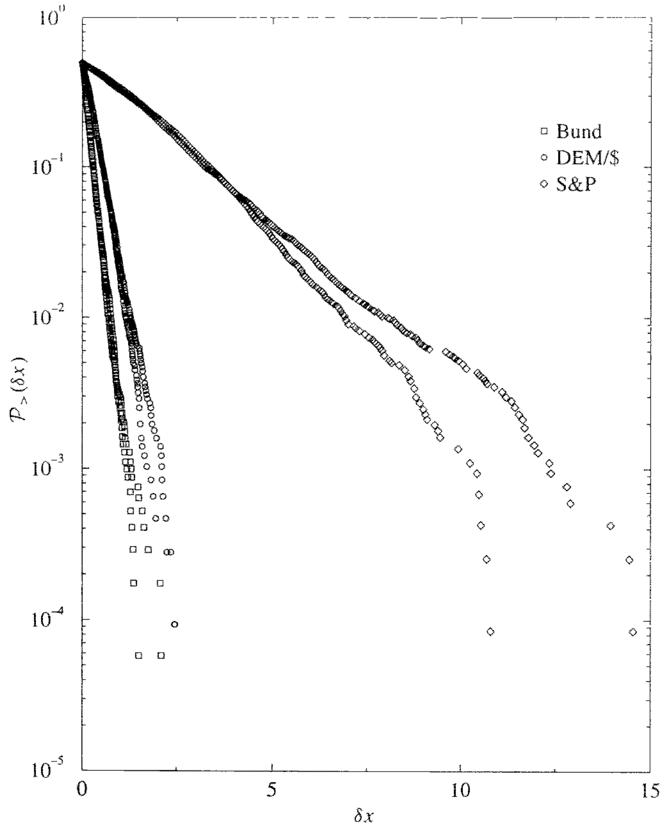

Fig. 2.12. Cumulative distribution of the price increments (positive and negative), on the scale of $N = 1$ day, for the three studied assets, and in a linear–log representation. One clearly see the approximate exponential nature of the tails, which are straight lines in this representation.

2.3.2 *Multiscaling – Hurst exponent (*)*

The fact that the autocorrelation function is zero beyond a certain time scale τ implies that the quantity $\langle[x_N - x_0]^2\rangle$ grows as $D\tau N$. However, measuring the temporal fluctuations using solely this quantity can be misleading. Take for example the case where the price increments δx_k are independent, but distributed according to a TLD of index $\mu < 2$. As we have explained in Section 1.6.5, the sum $x_N - x_0 = \sum_{k=0}^{N-1} \delta x_k$ behaves, as long as $N \ll N^* = \kappa$, as a pure 'Lévy' sum, for which the truncation is inessential. Its order of magnitude is therefore $x_N - x_0 \sim AN^{1/\mu}$, where A^μ is the tail parameter of the Lévy distribution. However, the second moment $\langle[x_N - x_0]^2\rangle = D\tau N$, is dominated by extreme fluctuations, and is therefore governed by the existence of the exponential truncation which gives a finite value to $D\tau$ proportional to $A^\mu\alpha^{\mu-2}$. One can check that as long as $N \ll N^*$, one has $\sqrt{D\tau N} \gg AN^{1/\mu}$. This means that in this case, the second moment overestimates the amplitude of probable fluctuations. One can generalize the above result to the qth moment of the price difference, $\langle[x_N - x_0]^q\rangle$. If $q > \mu$, one finds that *all moments* grow like N in the regime $N \ll N^*$, and like $N^{q/\mu}$ if $q < \mu$. This is to be contrasted with the sum of Gaussian variables, where $\langle[x_N - x_0]^q\rangle$ grows as $N^{q/2}$ for all $q > 0$. More generally, one can define an exponent ζ_q as $\langle[x_N - x_0]^q\rangle \propto N^{\zeta_q}$. If ζ_q/q is not constant with q, one speaks of *multiscaling*. It is not always easy to distinguish true multiscaling from apparent multiscaling, induced by crossover or finite size effects. For example, in the case where one sums uncorrelated random variables with a long-range correlation in the variance, one finds that the kurtosis decays slowly, as $\kappa_N \propto N^{-\nu}$, where $\nu < 1$ is the exponent governing the decay of the correlations. This means that the fourth moment of the difference $x_N - x_0$ behaves as:

$$\langle[x_N - x_0]^4\rangle = (D\tau N)^2[3 + \kappa_N] \sim N^2 + N^{2-\nu}. \tag{2.12}$$

If ν is small, one can fit the above expression, over a finite range of N, using an *effective* exponent $\zeta_4 < 2$, suggesting multiscaling. Similarly, higher moments can be accurately fitted using an effective exponent $\zeta_q < q/2$.[10] This is certainly a possibility that one should keep in mind, in particular when analysing financial time series (see Mandelbrot, 1998).

Another interesting way to characterize the temporal development of the fluctuations is to study, as suggested by Hurst, the average distance between the 'high' and the 'low' in a window of size $t = n\tau$:

$$\mathcal{H}(n) = \langle \max(x_k)_{k=\ell+1,\ell+n} - \min(x_k)_{k=\ell+1,\ell+n}\rangle_\ell. \tag{2.13}$$

The Hurst exponent H is defined from $\mathcal{H}(n) \propto n^H$. In the case where the increments δx_k are Gaussian, one finds that $H \equiv \frac{1}{2}$ (for large n). In the case of a TLD of index $1 < \mu \leq 2$, one finds:

$$\mathcal{H}(n) \propto An^{\frac{1}{\mu}} \quad (n \ll N^*) \qquad \mathcal{H}(n) \propto \sqrt{D\tau n} \quad (n \gg N^*). \tag{2.14}$$

The Hurst exponent therefore evolves from an anomalously high value $H = 1/\mu$ to the 'normal' value $H = \frac{1}{2}$ as n increases. Figure 2.13 shows the Hurst function $\mathcal{H}(n)$ for the three liquid assets studied here. One clearly sees that the 'local' exponent H slowly decreases from a high value ($\sim$0.7, quite close to $1/\mu = \frac{2}{3}$) at small times, to $H \simeq \frac{1}{2}$ at long times.

[10] For more details on this point, see: J.-P. Bouchaud, M. Potters, M. Meyer, Apparent multifractality in financial time series, *European Journal of Physics*, **13**, 595 (2000).

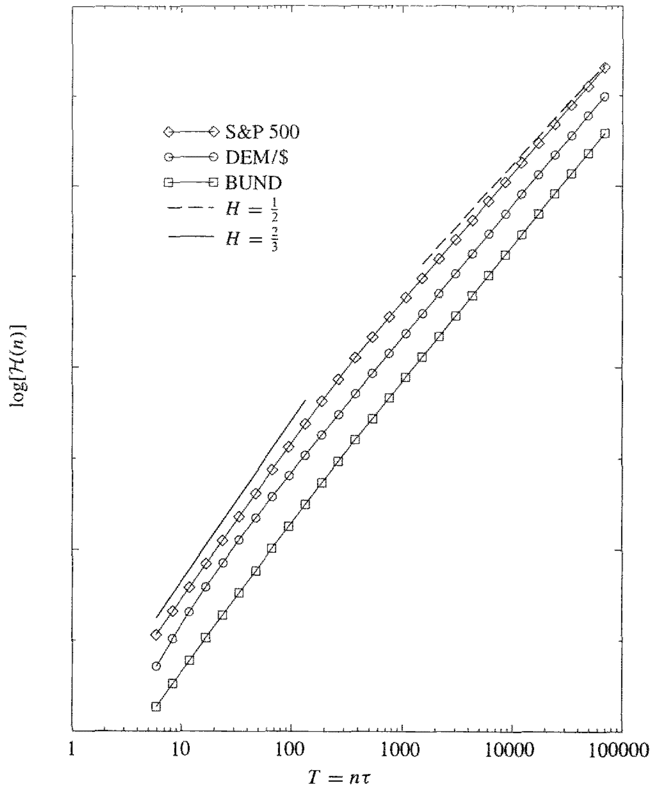

Fig. 2.13. Hurst Function $\mathcal{H}(n)$ (up to an arbitrary scaling factor) for the three liquid assets studied in this chapter, in log–log coordinates. The local slope gives the value of the Hurst exponent H. One clearly sees that this exponent goes from rather a high value for small n to a value close to $\frac{1}{2}$ when n increases.

2.4 Anomalous kurtosis and scale fluctuations

For in a minute there are many days.

(Shakespeare, *Romeo and Juliet.*)

As mentioned above, one sees in Figure 2.11 that $P(\delta x, N)$ systematically deviates from $[P_1(\delta x_1)]^{*N}$. In particular, the tails of $P(\delta x, N)$ are anomalously 'fat'. Equivalently, the kurtosis κ_N of $P(\delta x, N)$ is higher than κ_1/N, as one can see from Figure 2.14, where κ_N is plotted as a function of N in log–log coordinates.

Correspondingly, more complex correlations functions, such as that of the *squares* of δx_k, reveal a non-trivial behaviour. An interesting quantity to consider

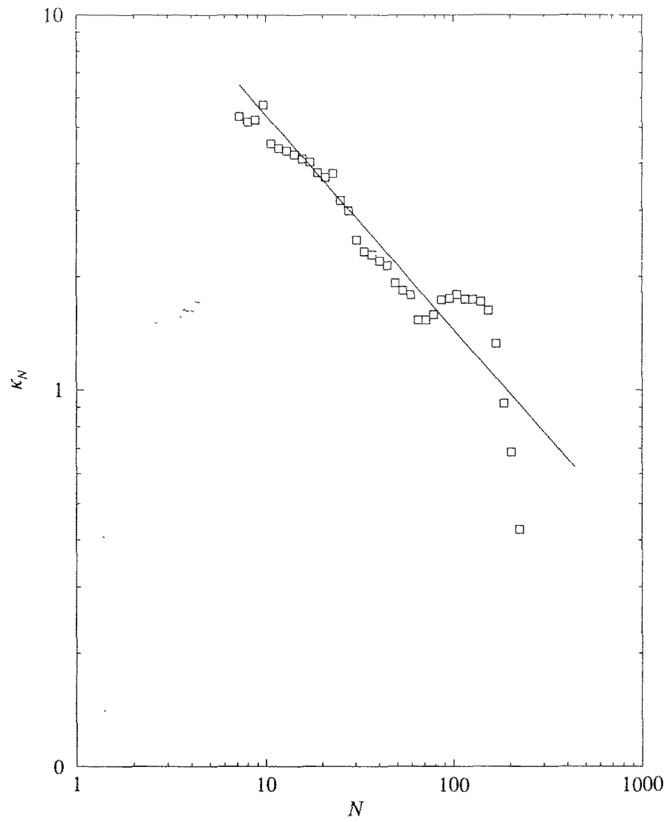

Fig. 2.14. Kurtosis κ_N at scale N as a function of N, for the Bund. In this case, the elementary time scale τ is 30 min. If the iid assumption were true, one should find $\kappa_N = \kappa_1/N$. The straight line has a slope of -0.43, which means that the decay of the kurtosis κ_N is much smaller, as $\simeq 20/N^{0.43}$.

is the amplitude of the fluctuations, averaged over one day, defined as:

$$\gamma = \frac{1}{N_d}\sum_{k=1}^{N_d} |\delta x_k|, \tag{2.15}$$

where δx_k is the 5-min increment, and N_d is the number of 5-min intervals within a day. This quantity is clearly strongly correlated in time (Figs 2.15 and 2.16): the periods of strong volatility persist far beyond the day time scale.

A simple way to account for these effects is to assume that the elementary distribution P_1 also depends of time. One actually observes that the level of activity on a market (measured by the volume of transactions) on a given time interval

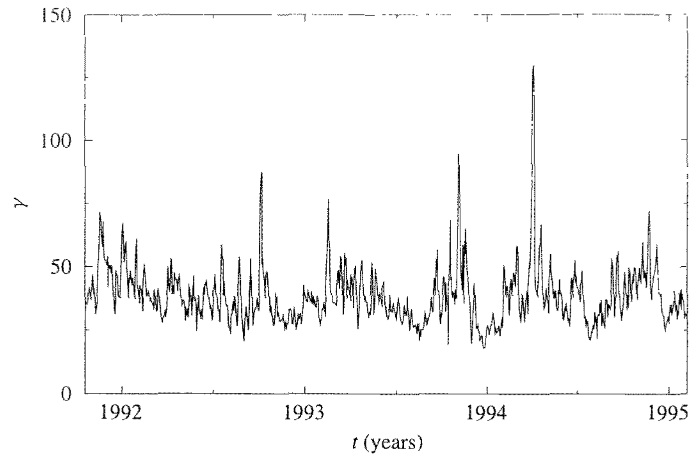

Fig. 2.15. Evolution of the 'volatility' γ as a function of time, for the S&P 500 in the period 1991–95. One clearly sees periods of large volatility, which persist in time.

can vary quite strongly with time. It is reasonable to think that the *scale* of the fluctuations γ of the price depends directly on the frequency and volume of the transactions. A simple hypothesis is that apart from a change of this level of activity, the mechanisms leading to a change of price are the same, and therefore that the fluctuations have the same distributions, up to a change of scale. More precisely, we shall assume that the distribution of price changes is such that:

$$P_{1k}(\delta x_k) = \frac{1}{\gamma_k} P_{10}\left(\frac{\delta x_k}{\gamma_k}\right), \tag{2.16}$$

where $P_{10}(u)$ is a certain distribution normalized to 1 and independent of k. The factor γ_k represents the scale of the fluctuations: one can define γ_k and P_{10} such that $\int |u| P_{10}(u)\, du \equiv 1$. The variance $D_k\tau$ is then proportional to γ_k^2.

In the case where P_{10} is Gaussian and in the limit of continuous time, the model defined by Eq. (2.16) is known in the literature as a 'stochastic volatility' model. The model defined by Eq. (2.16) is however more general since P_{10} is *a priori* arbitrary.

If one assumes that the random variables $\delta x_k/\gamma_k$ are independent and of zero mean, one can show (see Section 1.7.2 and Appendix A) that the average kurtosis of the distribution $P(\delta x, N)$ is given, for $N \geq 1$, by:

$$\kappa_N = \frac{1}{N}\left[\kappa_0 + (3+\kappa_0)g(0) + 6\sum_{\ell=1}^{N}\left(1 - \frac{\ell}{N}\right)g(\ell)\right], \tag{2.17}$$

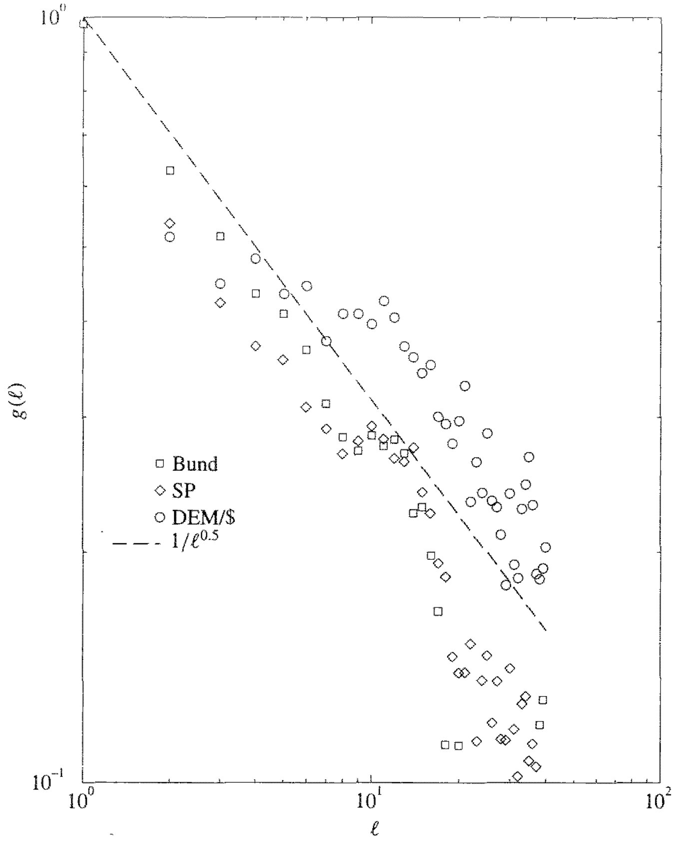

Fig. 2.16. Temporal correlation function, $\langle\gamma_k\gamma_{k+\ell}\rangle - \langle\gamma_k\rangle^2$, normalized by $\langle\gamma_k^2\rangle - \langle\gamma_k\rangle^2$. The value $\ell = 1$ corresponds to an interval of 1 day. For comparison, we have shown a decay as $1/\sqrt{\ell}$.

where κ_0 is the kurtosis of the distribution P_{10} defined above (cf. Eq. (2.16)), and g the correlation function of the variance $D_k\tau$ (or of the γ_k^2):

$$\overline{(D_k - \overline{D})(D_\ell - \overline{D})} = \overline{D}^2 g(|\ell - k|). \tag{2.18}$$

The overbar means that one should average over the fluctuations of the D_k. It is interesting to notice that even in the absence of 'bare' kurtosis ($\kappa_0 = 0$), volatility fluctuations are enough to induce a non-zero kurtosis $\kappa_1 = \kappa_0 + (3+\kappa_0)g(0)$.

The empirical data on the kurtosis are well accounted for using the above formula, with the choice $g(\ell) \propto \ell^{-\nu}$, with $\nu = 0.43$ in the case of the Bund. This choice for $g(\ell)$ is also in qualitative agreement with the decay of the correlations

of the γ's (Fig. 2.16). However, a fit of the data using for $g(\ell)$ the sum of two exponentials $\exp(-\ell/\ell_{1,2})$ is also acceptable. One finds two rather different time scales: ℓ_1 is shorter than a day, and a long correlation time ℓ_2 of a few tens of days.

One can thus quite clearly see that the scale of the fluctuations (known in the market as the volatility) changes with time, with a rather long persistence time scale. This slow evolution of the volatility in turn leads to an anomalous decay of the kurtosis κ_N as a function of N. As we shall see in Section 4.3.4, this has direct consequences for the dynamics of the volatility smile observed on option markets.

2.5 Volatile markets and volatility markets

We have considered, up to now, very liquid markets, where extreme price fluctuations are rather rare. On less liquid/less mature markets, the probability of extreme moves is much larger. The case of short-term interest rates is also interesting, since the evolution of, say, the 3-month rate is directly affected by the decision of central banks to increase or to decrease the day to day rate. As discussed further in Section 2.6 below, this leads to a rather high kurtosis, related to the fact that the short rate often does not change at all, but sometimes changes a lot. The kurtosis of the US 3-month rate is on the order of 20 for daily moves (Fig. 2.17). Emerging markets (such as South America or Eastern Europe markets) are obviously even wilder. The example of the Mexican peso (MXP) is interesting, because the cumulative distribution of the daily changes of the rate MXP/\$ reveals *power-law* tails, with no obvious truncation, with an exponent $\mu = 1.5$ (Fig. 2.18). This data-set corresponds to the years 1992–94, just *before* the crash of the peso (December 1994). A similar value of μ has also been observed, for example, in the fluctuations of the Budapest Stock Exchange.[11]

Another interesting quantity is the volatility itself which varies with time, as emphasized above. The price of options reflect quite accurately the value of the historical volatility in a recent past (see Section 4.3.4). Therefore, the volatility can be considered as a special type of asset, which one can study as such. We shall define as above the volatility γ_k as the average over a day of the absolute value of the 5-min price increments. The autocorrelation function of the γ's is shown in Figure 2.16; it is found to decrease slowly, perhaps as a power-law with an exponent ν in the range 0.1 to 0.5 (Fig. 2.16).[12] The distribution of the

[11] J. Rotyis, G. Vattay, Statistical analysis of the stock index of the Budapest Stock Exchange, in [Kondor and Kertecz].

[12] On this point, see, e.g. [Ding, Arneodo], and Y. Liu *et al.*, The statistical properties of volatility of price fluctuations, *Physical Review*, **E 60**, 1390 (1999).

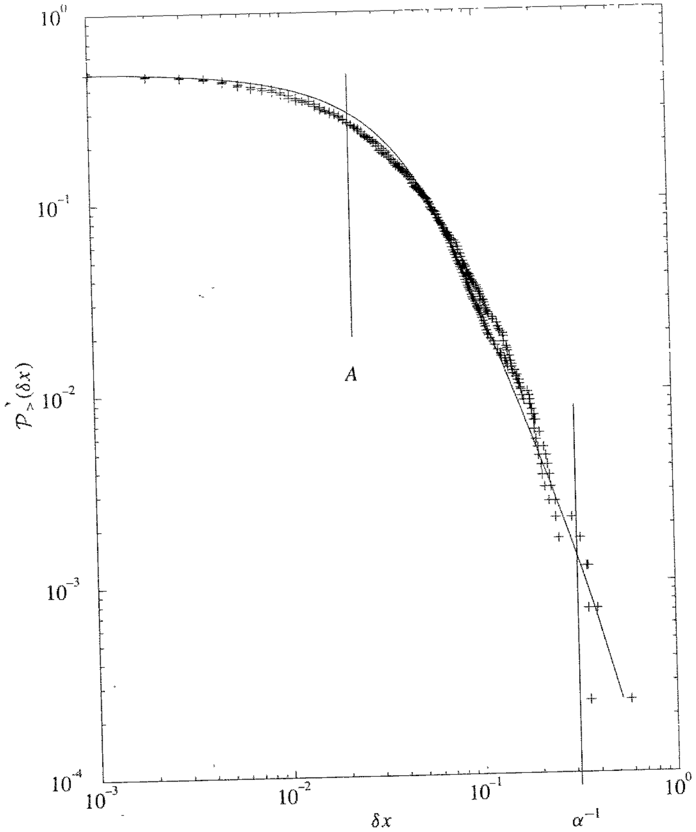

Fig. 2.17. Cumulative distribution $\mathcal{P}_{1>}(\delta x)$ (for $\delta x > 0$) and $\mathcal{P}_{1<}(\delta x)$ (for $\delta x < 0$), for the US 3-month rate (US T-Bills from 1987 to 1996), with $\tau = 1$ day. The thick line corresponds to the best fit using a symmetric TLD $L_\mu^{(t)}$, of index $\mu = \frac{3}{2}$. We have also shown the corresponding values of A and α^{-1}, which gives a kurtosis equal to 22.6.

measured volatility γ is shown in Figure 2.19 for the S&P 500, but other assets lead to similar curves. This distribution decreases slowly for large γ's, again as an exponential or a high power-law. Several functional forms have been suggested, such as a log-normal distribution, or an inverse Gamma distribution (see Section 2.9 for a specific model for this behaviour). However, one must keep in mind that the quantity γ is only an approximation for the 'true' volatility. The distribution shown in Figure 2.19 is therefore the convolution of the true distribution with a measurement error distribution.

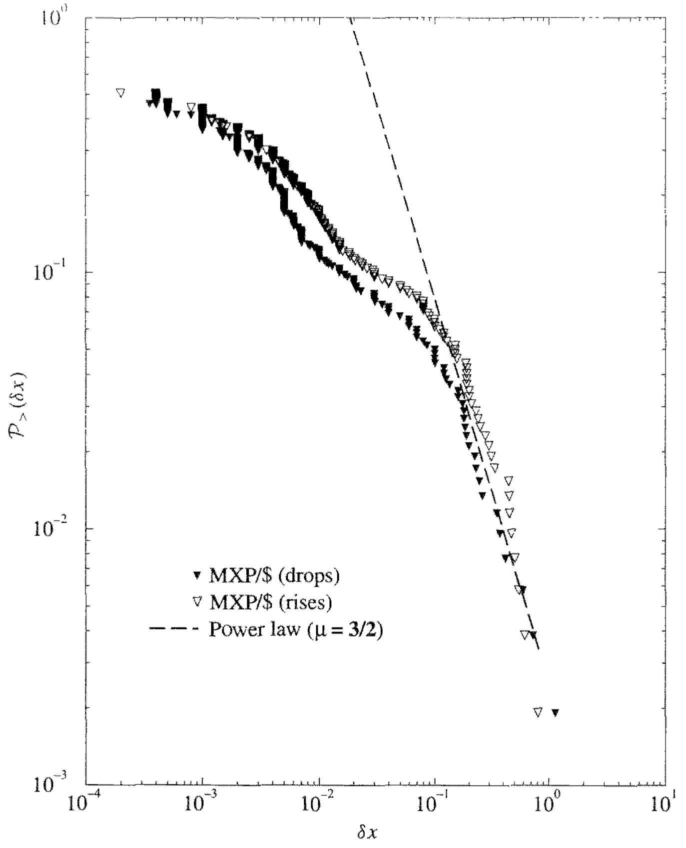

Fig. 2.18. Cumulative distribution $\mathcal{P}_{1>}(\delta x)$ (for $\delta x > 0$) and $\mathcal{P}_{1<}(\delta x)$ (for $\delta x < 0$), for the Mexican peso versus \$, with $\tau = 1$ day. The data corresponds to the years 1992–94. The thick line shows a power-law decay, with a value of $\mu = \frac{3}{2}$. The extrapolation to 10 years gives a most probable worst day on the order of -40%.

2.6 Statistical analysis of the forward rate curve (*)

The case of the interest rate curve is particularly complex and interesting, since it is not the random motion of a point, but rather the consistent history of a whole curve (corresponding to different loan maturities) which is at stake. The need for a consistent description of the whole interest rate curve is furthermore enhanced by the rapid development of interest rate derivatives (options, swaps,[13] options on swaps, etc.) [Hull].

[13] A swap is a contract where one exchanges fixed interest rate payments with floating interest rate payments.

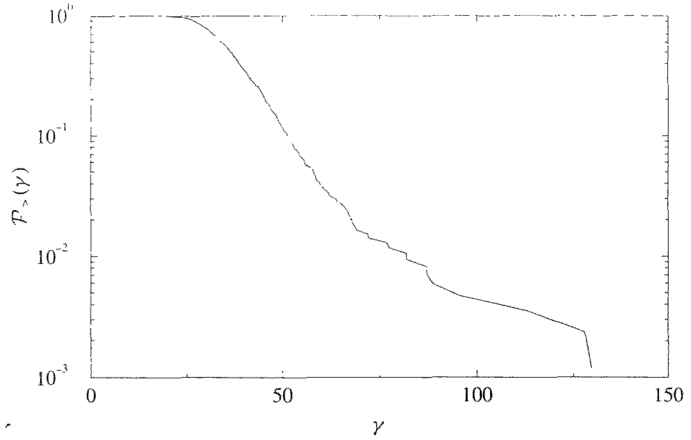

Fig. 2.19. Cumulative distribution of the measured volatility γ of the S&P, $\mathcal{P}_{1>}(\gamma)$, in a linear–log plot. Note that the tail of this distribution decays more slowly than exponentially.

Present models of the interest rate curve fall into two categories: the first one is the Vasicek model and its variants, which focuses on the dynamics of the short-term interest rate, from which the whole curve is reconstructed.[14] The second one, initiated by Heath, Jarrow and Morton takes the full forward rate curve as dynamic variables, driven by (one or several) continuous-time Brownian motion, multiplied by a maturity-dependent scale factor. Most models are however primarily motivated by their mathematical tractability rather than by their ability to describe the data. For example, the fluctuations are often assumed to be Gaussian, thereby neglecting 'fat tail' effects.

Our aim in this section is not to discuss these models in any detail, but rather to present an empirical study of the forward rate curve (FRC), where we isolate several important qualitative features that a good model should be asked to reproduce.[15] Some intuitive arguments are proposed to relate the series of observations reported below.

2.6.1 Presentation of the data and notations

The forward interest rate curve (FRC) at time t is fully specified by the collection of all *forward rates* $f(t, \theta)$, for different maturities θ. It allows us for example to calculate the price $B(t, \theta)$ at time t of a (so-called 'zero-coupon') bond, which by

[14] For a compilation of the most important theoretical papers on the interest rate curve, see: [Hughston].

[15] This section is based on the following papers: J.-P. Bouchaud, N. Sagna, R. Cont, N. ElKaroui, M. Potters, Phenomenology of the interest rate curve. *Applied Mathematical Finance*, **6**, 209 (1999) and *idem*, Strings attached, *Risk Magazine*, **11** (7), 56 (1998).

definition pays 1 at time $t+\theta$. The forward rates are by definition such that they compound to give $B(t,\theta)$:

$$B(t,\theta) = \exp\left(-\int_0^\theta f(t,u)\,\mathrm{d}u\right), \tag{2.19}$$

$r(t) = f(t,\theta=0)$ is called the 'spot rate'. Note that in the following θ is always a time difference; the maturity *date* T is $t+\theta$.

Our study is based on a data-set of daily prices of Eurodollar futures contracts on interest rates.[16] The interest rate underlying the Eurodollar futures contract is a 90-day rate, earned on dollars deposited in a bank outside the US by another bank. The interest in studying forward rates rather than yield curves is that one has a direct access to a 'derivative' (in the mathematical sense: $f(t,\theta) = -\partial \log B(t,\theta)/\partial\theta$), which obviously contains more precise information than the yield curve (defined from the logarithm of $B(t,\theta)$) itself.

In practice, the futures markets price 3-months forward rates for *fixed* expiration dates, separated by 3-month intervals. Identifying 3-months futures rates with instantaneous forward rates, we have available a sequence of time series on forward rates $f(t, T_i - t)$, where T_i are fixed dates (March, June, September and December of each year). We can convert these into fixed maturity (multiple of 3-months) forward rates by a simple linear interpolation between the two nearest points such that $T_i - t \le \theta \le T_{i+1} - t$. Between 1990 and 1996, one has at least 15 different Eurodollar maturities for each market date. Between 1994 and 1996, the number of available maturities rises to 30 (as time grows, longer and longer maturity forward rates are being traded on future markets); we shall thus often use this restricted data-set. Since we only have daily data, our reference time scale will be $\tau = 1$ day. The variation of $f(t,\theta)$ between t and $t+\tau$ will be denoted as $\delta f(t,\theta)$:

$$\delta f(t,\theta) = f(t+\tau,\theta) - f(t,\theta). \tag{2.20}$$

2.6.2 *Quantities of interest and data analysis*

The description of the FRC has two, possibly interrelated, aspects:

(i) What is, at a given instant of time, the *shape* of the FRC as a function of the maturity θ?

(ii) What are the statistical properties of the increments $\delta f(t,\theta)$ between time t and time $t+\tau$, and how are they correlated with the shape of the FRC at time t?

[16] In principle forward contracts and futures contracts are not strictly identical – they have different margin requirements – and one may expect slight differences, which we shall neglect in the following.

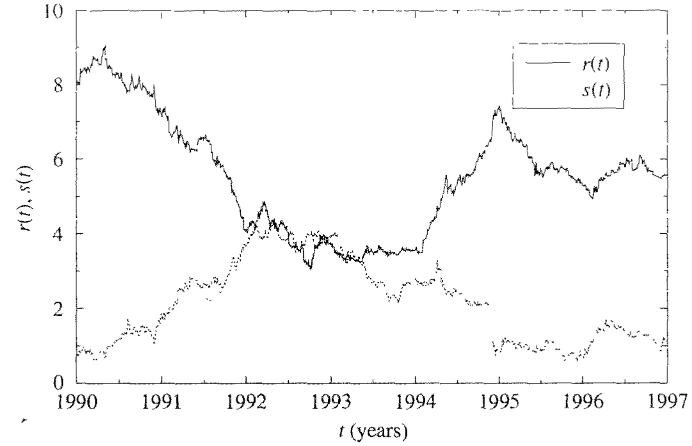

Fig. 2.20. The historical time series of the spot rate $r(t)$ from 1990 to 1996 (top curve) – actually corresponding to a 3-month future rate (dark line) and of the 'spread' $s(t)$ (bottom curve), defined with the longest maturity available over the whole period 1990–96 on future markets, i.e. $\theta_{\max} = 4$ years.

The two basic quantities describing the FRC at time t are the value of the short-term interest rate $f(t,\theta_{\min})$ (where $\theta_{\min}$ is the shortest available maturity), and that of the short-term/long-term *spread* $s(t) = f(t,\theta_{\max}) - f(t,\theta_{\min})$, where $\theta_{\max}$ is the longest available maturity. The two quantities $r(t) \simeq f(t,\theta_{\min})$, $s(t)$ are plotted versus time in Figure 2.20;[17] note that:

- The volatility σ of the spot rate $r(t)$ is equal to $0.8\%/\sqrt{\text{year}}$.[18] This obtained by averaging over the whole period.
- The spread $s(t)$ has varied between 0.53 and 4.34%. Contrarily to some European interest rates on the same period, $s(t)$ has always remained positive. (This however does not mean that the FRC is increasing monotonically, see below.)

Figure 2.21 shows the average shape of the FRC, determined by averaging the difference $f(t,\theta) - r(t)$ over time. Interestingly, this function is rather well fitted by a simple square-root law. This means that on average, the difference between the forward rate with maturity θ and the spot rate is equal to $a\sqrt{\theta}$, with a proportionality constant $a = 0.85\%/\sqrt{\text{year}}$ which turns out to be nearly identical to the spot rate volatility. We all propose a simple interpretation of this fact below.

[17] We shall from now on take the 3-month rate as an approximation to the spot rate $r(t)$.

[18] The dimension of r should really be % per year, but we conform here to the habit of quoting r simply in %. Note that this can sometimes be confusing when checking the correct dimensions of a formula.

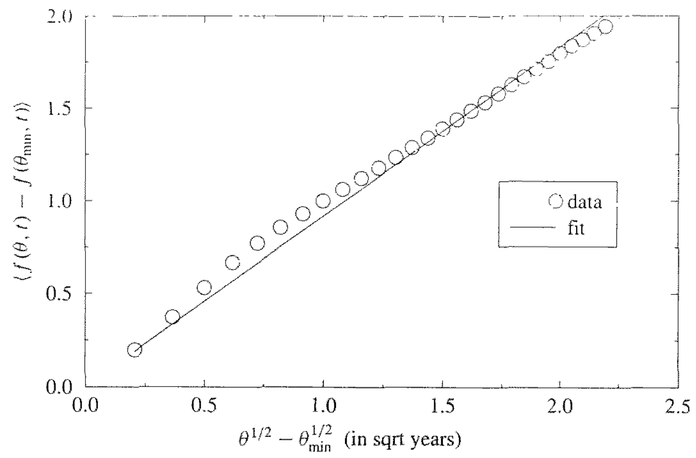

Fig. 2.21. The average FRC in the period 1994–96, as a function of the maturity θ. We have shown for comparison a *one parameter fit* with a square-root law, $a(\sqrt{\theta} - \sqrt{\theta_{min}})$. The same $\sqrt{\theta}$ behaviour actually extends up to $\theta_{max} = 10$ years, which is available in the second half of the time period.

Let us now turn to an analysis of the *fluctuations* around the average shape. These fluctuations are actually similar to that of a vibrating elastic string. The average deviation $\Delta(\theta)$ can be defined as:

$$\Delta^2(\theta) \equiv \left\langle \left[f(t,\theta) - r(t) - s(t)\sqrt{\frac{\theta}{\theta_{max}}} \right]^2 \right\rangle, \tag{2.21}$$

and is plotted in Figure 2.22, for the period 1994–96. The maximum of Δ is reached for a maturity of $\theta^* = 1$ year.

We now turn to the statistics of the *daily* increments $\delta f(t,\theta)$ of the forward rates, by calculating their volatility $\sigma(\theta) = \sqrt{\langle \delta f(t,\theta)^2 \rangle}$ and their excess kurtosis

$$\kappa(\theta) = \frac{\langle \delta f(t,\theta)^4 \rangle}{\sigma^4(\theta)} - 3. \tag{2.22}$$

A very important quantity will turn out to be the following 'spread' correlation function:

$$C(\theta) = \frac{\langle \delta f(t,\theta_{min})\,(\delta f(t,\theta) - \delta f(t,\theta_{min}))\rangle}{\sigma^2(\theta_{min})}, \tag{2.23}$$

which measures the influence of the short-term interest fluctuations on the other modes of motion of the FRC, subtracting away any trivial overall translation of the FRC.

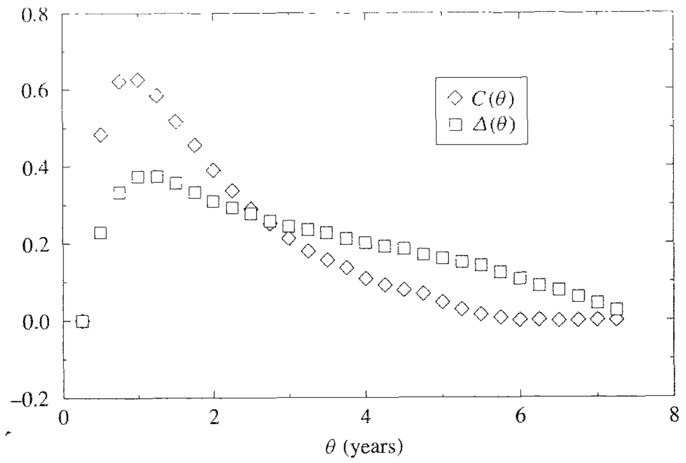

Fig. 2.22. Root mean square deviation $\Delta(\theta)$ from the average FRC as a function of θ. Note the maximum for $\theta^* = 1$ year, for which $\Delta \simeq 0.38\%$. We have also plotted the correlation function $C(\theta)$ (defined by Eq. (2.23)) between the daily variation of the spot rate and that of the forward rate at maturity θ, in the period 1994–96. Again, $C(\theta)$ is maximum for $\theta = \theta^*$, and decays rapidly beyond.

Figure 2.23 shows $\sigma(\theta)$ and $\kappa(\theta)$. Somewhat surprisingly, $\sigma(\theta)$, much like $\Delta(\theta)$ has a maximum around $\theta^* = 1$ year. The order of magnitude of $\sigma(\theta)$ is $0.05\%/\sqrt{\text{day}}$, or $0.8\%/\sqrt{\text{year}}$. The daily kurtosis $\kappa(\theta)$ is rather high (on the order of 5), and only weakly decreasing with θ.

Finally, $C(\theta)$ is shown in Figure 2.22; its shape is again very similar to those of $\Delta(\theta)$ and $\sigma(\theta)$, with a pronounced maximum around $\theta^* = 1$ year. This means that the fluctuations of the short-term rate are *amplified* for maturities around 1 year. We shall come back to this important point below.

2.6.3 Comparison with the Vasicek model

The simplest FRC model is a one-factor model due to Vasicek, where the whole term structure can be ascribed to the short-term interest rate. The latter is assumed to follow a so-called 'Ornstein–Uhlenbeck' (or mean reverting) process defined as:

$$\frac{dr(t)}{dt} = \Omega(r_0 - r(t)) + \sigma\xi(t), \tag{2.24}$$

where r_0 is an 'equilibrium' reference rate, Ω describes the strength of the reversion towards r_0 (and is the inverse of the mean reversion time), and $\xi(t)$ is a Gaussian noise, of volatility 1. In its simplest version, the Vasicek model prices

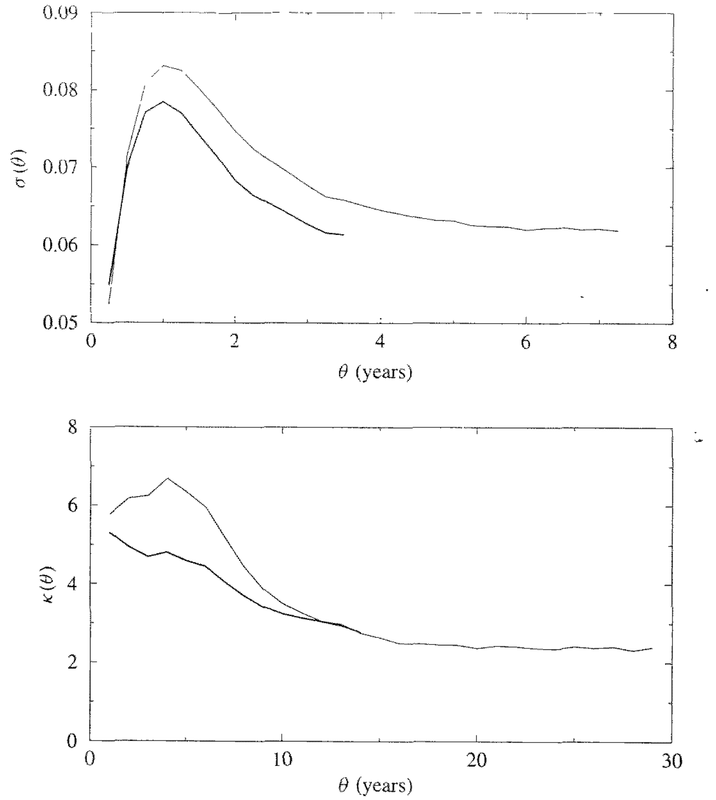

Fig. 2.23. The daily volatility and kurtosis as a function of maturity. Note the maximum of the volatility for $\theta = \theta^*$, while the kurtosis is rather high, and only very slowly decreasing with θ. The two curves correspond to the periods 1990–96 and 1994–96, the latter period extending to longer maturities.

a bond maturing at T as the following average:

$$B(t, T) = \left\langle \exp - \int_t^T r(u)\,\mathrm{d}u \right\rangle, \tag{2.25}$$

where the averaging is over the possible histories of the spot rate between now and the maturity, where the uncertainty is modelled by the noise ξ. The computation of the above average is straightforward when ξ is Gaussian, and leads to (using Eq. (2.19)):

$$f(t, \theta) = r(t) + (r_0 - r(t))(1 - \mathrm{e}^{-\Omega\theta}) - \frac{\sigma^2}{2\Omega^2}(1 - \mathrm{e}^{-\Omega\theta})^2. \tag{2.26}$$

The basic results of this model are as follows:

- Since $\langle r_0 - r(t)\rangle = 0$, the average of $f(t, \theta) - r(t)$ is given by

$$\langle f(t, \theta) - r(t)\rangle = -\sigma^2/2\Omega^2(1 - \mathrm{e}^{-\Omega\theta})^2. \tag{2.27}$$

  and should thus be *negative*, at variance with empirical data. Note that in the limit $\Omega\theta \ll 1$, the order of magnitude of this (negative) term is very small: taking $\sigma = 1\%/\sqrt{\text{year}}$ and $\theta = 1$ year, it is found to be equal to 0.005%, much smaller than the typical differences actually observed on forward rates.
- The volatility $\sigma(\theta)$ is monotonically decreasing as $\exp -\Omega\theta$, while the kurtosis $\kappa(\theta)$ is identically zero (because ξ is Gaussian).
- The correlation function $\mathcal{C}(\theta)$ is negative and is a monotonic decreasing function of its argument, in total disagreement with observations (Fig. 2.22).
- The variation of the spread $s(t)$ and of the spot rate should be perfectly correlated, which is not the case (Fig. 2.22): more than one factor is in any case needed to account for the deformation of the FRC.

An interesting extension of Vasicek's model designed to fit exactly the 'initial' FRC $f(t = 0, \theta)$ was proposed by Hull and White [Hull]. It amounts to replacing the above constants Ω and r_0 by time-dependent functions. For example, $r_0(t)$ represents the anticipated evolution of the 'reference' short-term rate itself with time. These functions can be adjusted to fit $f(t = 0, \theta)$ exactly. Interestingly, one can then derive the following relation:

$$\left\langle \frac{\partial r(t)}{\partial t} \right\rangle = \left\langle \frac{\partial f}{\partial \theta}(t, 0) \right\rangle, \tag{2.28}$$

up to a term of order σ^2 which turns out to be negligible, exactly for the same reason as explained above. On average, the second term (estimated by taking a finite difference estimate of the partial derivative using the first two points of the FRC) is definitely found to be positive, and equal to 0.8%/year. On the same period (1990–96), however, the spot rate has decreased from 8.1 to 5.9%, instead of growing by $7 \times 0.8\% = 5.6\%$.

In simple terms, both the Vasicek and the Hull–White model mean the following: the FRC should basically reflect the market's expectation of the average evolution of the spot rate (up to a correction of the order of σ^2, but which turns out to be very small, see above). However, since the FRC is on average increasing with the maturity (situations when the FRC is 'inverted' are comparatively much rarer), this would mean that the market systematically expects the spot rate to rise, which it does not. It is hard to believe that the market persists in error for such a long time. Hence, the upward slope of the FRC is not only related to what the market expects on average, but that a systematic risk premium is needed to account for this increase.

2.6.4 Risk-premium and the $\sqrt{\theta}$ law

The average FRC and value-at-risk pricing

The observation that on average the FRC follows a simple $\sqrt{\theta}$ law (i.e. $\langle f(t,\theta) - r(t)\rangle \propto \sqrt{\theta}$) suggests an intuitive, direct interpretation. At any time t, the market anticipates either a future rise, or a decrease of the spot rate. However, the average anticipated trend is, in the long run, zero, since the spot rate has bounded fluctuations. Hence, the *average* market's expectation is that the future spot rate $r(t)$ will be close to its present value $r(t = 0)$. In this sense, the average FRC should thus be flat. However, even in the absence of any trend in the spot rate, its *probable* change between now and $t = \theta$ is (assuming the simplest random walk behaviour) of the order of $\sigma\sqrt{\theta}$, where σ is the volatility of the spot rate. Money lenders agree at time t on a loan at rate $f(t,\theta)$, which will run between time $t+\theta$ and $t+\theta+d\theta$. These money lenders will themselves borrow money from central banks at the short-term rate prevailing at that date, i.e. $r(t+\theta)$. They will therefore lose money whenever $r(t+\theta) > f(t,\theta)$. Hence, money lenders take a bet on the future value of the spot rate and want to be sure not to lose their bet more frequently than, say, once out of five. Thus their price for the forward rate is such that the probability that the spot rate at time $t+\theta$, $r(t+\theta)$ actually exceeds $f(t,\theta)$ is equal to a certain number p:

$$\int_{f(t,\theta)}^{\infty} P(r', t+\theta|r, t)\,\mathrm{d}r' = p, \tag{2.29}$$

where $P(r', t'|r, t)$ is the probability that the spot rate is equal to r' at time t' knowing that it is r now (at time t). Assuming that r' follows a simple random walk centred around $r(t)$ then leads to:[19]

$$f(t,\theta) = r(t) + a\sigma(0)\sqrt{\theta}, \qquad a = \sqrt{2}\,\mathrm{erfc}^{-1}(2p). \tag{2.30}$$

which indeed matches the empirical data, with $p \simeq 0.16$.

Hence, the shape of today's FRC can be thought of as an envelope for the probable future evolutions of the spot rate. The market appears to price future rates through a *Value at Risk* procedure (Eqs. (2.29) and (2.30) – see Chapter 3 below) rather than through an averaging procedure.

The anticipated trend and the volatility hump

Let us now discuss, along the same lines, the shape of the FRC at a given instant of time, which of course deviates from the average square root law. For a given instant of time t, the market actually expects the spot rate to perform a biased random walk. We shall argue that a consistent interpretation is that the market estimates the trend $m(t)$ by extrapolating the past behaviour of the spot rate itself. Hence, the probability distribution $P(r', t+\theta|r, t)$ used by the market is not centred around $r(t)$ but rather around:

$$r(t) + \int_t^{t+\theta} m(t, t+u)\,\mathrm{d}u, \tag{2.31}$$

where $m(t,t')$ can be called the anticipated bias at time t', seen from time t.

It is reasonable to think that the market estimates m by extrapolating the recent past to the nearby future. Mathematically, this reads:

$$m(t, t+u) = m_1(t)\mathcal{Z}(u) \text{ where } m_1(t) \equiv \int_0^{\infty} K(v)\delta r(t-v)\,\mathrm{d}v, \tag{2.32}$$

and where $K(v)$ is an averaging kernel of the past variations of the spot rate. One may call $\mathcal{Z}(u)$ the trend persistence function; it is normalized such that $\mathcal{Z}(0) = 1$, and describes how the present trend is expected to persist in the future. Equation (2.29) then gives:

$$f(t,\theta) = r(t) + A\sigma\sqrt{\theta} + m_1(t)\int_0^{\theta} \mathcal{Z}(u)\,\mathrm{d}u. \tag{2.33}$$

This mechanism is a possible explanation of why the three functions introduced above, namely $\Delta(\theta)$, $\sigma(\theta)$ and the correlation function $C(\theta)$ have similar shapes. Indeed, taking for simplicity an exponential averaging kernel $K(v)$ of the form $\epsilon\exp[-\epsilon v]$, one finds:

$$\frac{\mathrm{d}m_1(t)}{\mathrm{d}t} = -\epsilon m_1 + \epsilon\frac{\mathrm{d}r(t)}{\mathrm{d}t} + \epsilon\xi(t), \tag{2.34}$$

where $\xi(t)$ is an independent noise of strength σ_ξ^2, added to introduce some extra noise in the determination of the anticipated bias. In the absence of temporal correlations, one can compute from the above equation the average value of m_1^2. It is given by:

$$\langle m_1^2\rangle = \frac{\epsilon}{2}(\sigma^2(0) + \sigma_\xi^2). \tag{2.35}$$

In the simple model defined by Eq. (2.33) above, one finds that the correlation function $C(\theta)$ is given by:[20]

$$C(\theta) = \epsilon\int_0^{\theta} \mathcal{Z}(u)\,\mathrm{d}u. \tag{2.36}$$

Using the above result for $\langle m_1^2\rangle$, one also finds:

$$\Delta(\theta) = \sqrt{\frac{\sigma^2(0) + \sigma_\xi^2}{2\epsilon}}\,C(\theta), \tag{2.37}$$

thus showing that $\Delta(\theta)$ and $C(\theta)$ are in this model simply proportional.

Turning now to the volatility $\sigma(\theta)$, one finds that it is given by:

$$\sigma^2(\theta) = [1 + C(\theta)]^2\sigma^2(0) + C(\theta)^2\sigma_\xi^2. \tag{2.38}$$

We thus see that the maximum of $\sigma(\theta)$ is indeed related to that of $C(\theta)$. Intuitively, the reason for the volatility maximum is as follows: a variation in the spot rate changes that

[19] This assumption is certainly inadequate for small times, where large kurtosis effects are present. However, on the scale of months, these non-Gaussian effects can be considered as small.

[20] In reality, one should also take into account the fact that $a\sigma(0)$ can vary with time. This brings an extra contribution both to $C(\theta)$ and to $\sigma(\theta)$.

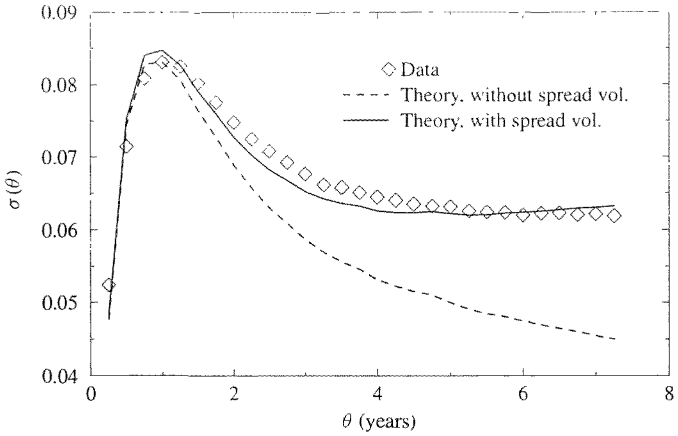

Fig. 2.24. Comparison between the theoretical prediction and the observed daily volatility of the forward rate at maturity θ, in the period 1994–96. The dotted line corresponds to Eq. (2.38) with $\sigma_\xi = \sigma(0)$, and the full line is obtained by adding the effect of the variation of the coefficient $a\sigma(0)$ in Eq. (2.33), which adds a contribution proportional to θ.

market anticipation for the trend $m_1(t)$. But this change of trend obviously has a larger effect when multiplied by a longer maturity. For maturities beyond 1 year, however, the decay of the persistence function comes into play and the volatility decreases again. The relation Eq. (2.38) is tested against real data in Figure 2.24. An important prediction of the model is that the deformation of the FRC should be strongly correlated with the past trend of the spot rate, averaged over a time scale $1/\epsilon$ (see Eq. (2.34)). This correlation has been convincingly established recently, with $1/\epsilon \simeq 100$ days.[21]

2.7 Correlation matrices (*)

As we shall see in Chapter 3, an important aspect of risk management is the estimation of the correlations between the price movements of different assets. The probability of large losses for a certain portfolio or option book is dominated by correlated moves of its different constituents – for example, a position which is simultaneously long in stocks and short in bonds will be risky because stocks and bonds move in opposite directions in crisis periods. The study of correlation (or covariance) matrices thus has a long history in finance, and is one of the cornerstone of Markowitz's theory of optimal portfolios (see Section 3.3). However, a reliable empirical determination of a correlation matrix turns out to be difficult: if one considers M assets, the correlation matrix contains $M(M-1)/2$ entries, which must be determined from M time series of length N; if N is not very large

[21] See: A. Matacz, J.-P. Bouchaud, An empirical study of the interest rate curve, to appear in *International Journal of Theoretical and Applied Finance* (2000).

compared to M, one should expect that the determination of the covariances is noisy, and therefore that the empirical correlation matrix is to a large extent random, i.e. the structure of the matrix is dominated by 'measurement' noise. If this is the case, one should be very careful when using this correlation matrix in applications. From this point of view, it is interesting to compare the properties of an empirical correlation matrix $\mathbf{C}$ to a 'null hypothesis' purely *random* matrix as one could obtain from a finite time series of strictly uncorrelated assets. Deviations from the random matrix case might then suggest the presence of true information.[22]

The empirical correlation matrix $\mathbf{C}$ is constructed from the time series of price changes δx_k^i (where i labels the asset and k the time) through the equation:

$$\mathbf{C}_{ij} = \frac{1}{N}\sum_{k=1}^{N} \delta x_k^i \delta x_k^j. \qquad (2.39)$$

In the following we assume that the average value of the δx's has been subtracted off, and that the δx's are rescaled to have a constant unit volatility. The null hypothesis of independent assets, which we consider now, translates itself in the assumption that the coefficients δx_k^i are independent, identically distributed, random variables.[23] The theory of random matrices, briefly expounded in Section 1.8, allows one to compute the density of eigenvalues of $\mathbf{C}$, $\rho_C(\lambda)$, in the limit of very large matrices: it is given by Eq. (1.120), with $Q = N/M$.

Now, we want to compare the empirical distribution of the eigenvalues of the correlation matrix of stocks corresponding to different markets with the theoretical prediction given by Eq. (1.120), based on the assumption that the correlation matrix is random. We have studied numerically the density of eigenvalues of the correlation matrix of $M = 406$ assets of the S&P 500, based on daily variations during the years 1991–96, for a total of $N = 1309$ days (the corresponding value of Q is 3.22). An immediate observation is that the highest eigenvalue λ_1 is 25 times larger than the predicted $\lambda_{\max}$ (Fig. 2.25, inset). The corresponding eigenvector is, as expected, the 'market' itself, i.e. it has roughly equal components on all the M stocks. The simplest 'pure noise' hypothesis is therefore inconsistent with the value of λ_1. A more reasonable idea is that the components of the correlation matrix which are orthogonal to the 'market' is pure noise. This amounts to subtracting the contribution of $\lambda_{\max}$ from the nominal value $\sigma^2 = 1$, leading to $\sigma^2 = 1 - \lambda_{\max}/M = 0.85$. The corresponding fit of the empirical distribution is shown as a dotted line in Figure 2.25. Several eigenvalues are still above $\lambda_{\max}$ and might contain some information, thereby reducing the variance of the effectively

[22] This section is based on the following paper: L. Laloux, P. Cizeau, J.-P. Bouchaud, M. Potters, Random matrix theory, *RISK Magazine*, **12**, 69 (March 1999).

[23] Note that even if the 'true' correlation matrix $\mathbf{C}_{\text{true}}$ is the identity matrix, its empirical determination from a finite time series will generate non-trivial eigenvectors and eigenvalues.

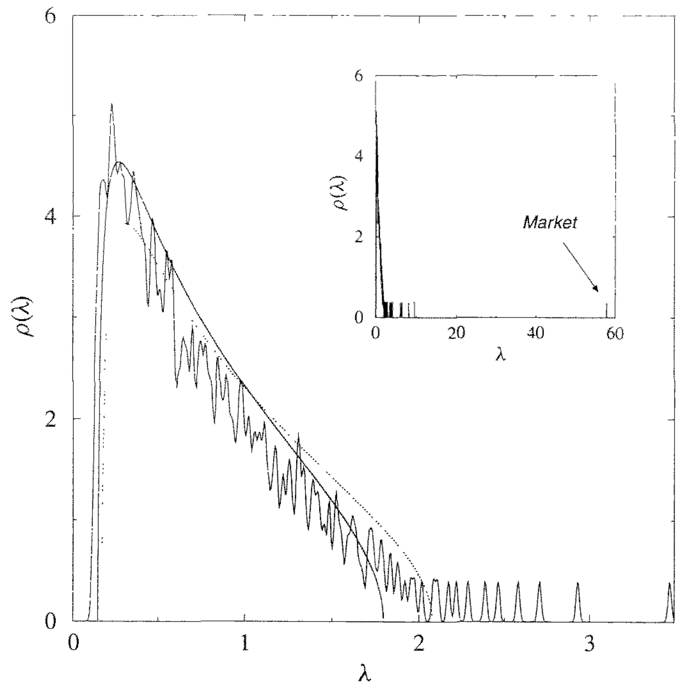

Fig. 2.25. Smoothed density of the eigenvalues of **C**, where the correlation matrix **C** is extracted from $M = 406$ assets of the S&P 500 during the years 1991–96. For comparison we have plotted the density Eq. (1.120) for $Q = 3.22$ and $\sigma^2 = 0.85$: this is the theoretical value obtained assuming that the matrix is purely random except for its highest eigenvalue (dotted line). A better fit can be obtained with a smaller value of $\sigma^2 = 0.74$ (solid line), corresponding to 74% of the total variance. Inset: same plot, but including the highest eigenvalue corresponding to the 'market', which is found to be ~ 30 times greater than $\lambda_{\max}$.

random part of the correlation matrix. One can therefore treat σ^2 as an adjustable parameter. The best fit is obtained for $\sigma^2 = 0.74$, and corresponds to the dark line in Figure 2.25, which accounts quite satisfactorily for 94% of the spectrum, whereas the 6% highest eigenvalues still exceed the theoretical upper edge by a substantial amount. These 6% highest eigenvalues are however responsible for 26% of the total volatility.

One can repeat the above analysis on different stock markets (e.g. Paris, London, Zurich), or on *volatility correlation matrices*, to find very similar results. In a first approximation, the location of the theoretical edge, determined by fitting the part of the density which contains most of the eigenvalues, allows one to distinguish 'information' from 'noise'.

The conclusion of this section is therefore that a large part of the empirical correlation matrices must be considered as 'noise', and cannot be trusted for risk management. In the next chapter, we will dwell on Markowitz' portfolio theory, which assumes that the correlation matrix is perfectly known. This theory must therefore be taken with a grain of salt, bearing in mind the results of the present section.

2.8 A simple mechanism for anomalous price statistics (*)

We have chosen the family of TLD to represent the distribution of price fluctuations. As mentioned above, Student distributions can also account quite well for the shape of the empirical distributions. Hyperbolic distributions have also been proposed. The choice of TLDs was motivated by two particular arguments:

- This family of distributions generalizes in a natural way the two classical descriptions of price fluctuations, since the Gaussian corresponds to $\mu = 2$, and the stable Lévy distributions correspond to $\alpha = 0$.
- The idea of TLD allows one to account for the deformation of the distributions as the time horizon N increases, and the anomalously high value of the Hurst exponent H at small times, crossing over to $H = \frac{1}{2}$ for longer times.

However, in order to justify the choice of one family of laws over the others, one needs a microscopic model for price fluctuations where a theoretical distribution can be computed. In the next two sections, we propose such 'models' (in the physicist's sense). These models are not very realistic, but only aim at showing that power-law distributions (possibly with an exponential truncation) appear quite naturally. Furthermore, the model considered in this section leads to a value of $\mu = \frac{3}{2}$, close to the one observed on real data.[24]

We assume that the price increment δx_k reflects the instantaneous offset between supply and demand. More precisely, if each operator on the market α wants to buy or sell a certain fixed quantity q of the asset X, one has:

$$\delta x_k \propto q \sum_\alpha \varphi_\alpha, \tag{2.40}$$

where φ_α can take the values -1, 0 or $+1$, depending on whether the operator α is selling, inactive, or buying. Suppose now that the operators interact among themselves in an heterogeneous manner: with a small probability $p/\mathcal{N}$ (where $\mathcal{N}$ is the total number of operators on the market), two operators α and β are

[24] This model was proposed in R. Cont, J.-P. Bouchaud, Herd behavior and aggregate fluctuations in financial markets, to appear in *Journal of Macroeconomic Dynamics* (1999). See also: D. Stauffer, P. M. C. de Olivera, A. T. Bernardes, Monte Carlo simulation of volatility clustering in market model with herding, *International Journal of Theoretical and Applied Finance* **2**, 83 (1999).

'connected', and with probability $1 - p/\mathcal{N}$, they ignore each other. The factor $1/\mathcal{N}$ means that on average, the number of operators connected to any particular one is equal to p. Suppose finally that if two operators are connected, they come to agree on the strategy they should follow, i.e. $\varphi_\alpha = \varphi_\beta$.

It is easy to understand that the population of operators clusters into groups sharing the same opinion. These clusters are defined such that there exists a connection between any two operators belonging to this cluster, although the connection can be indirect and follow a certain 'path' between operators. These clusters do not all have the same size, i.e. do not contain the same number of operators. If the size of cluster $\mathcal{A}$ is called $N(\mathcal{A})$, one can write:

$$\delta x_k \propto q \sum_{\mathcal{A}} N(\mathcal{A})\varphi(\mathcal{A}), \tag{2.41}$$

where $\varphi(\mathcal{A})$ is the common opinion of all operators belonging to $\mathcal{A}$. The statistics of the price increments δx_k therefore reduces to the statistics of the size of clusters, a classical problem in percolation theory [Stauffer]. One finds that as long as $p < 1$ (less than one 'neighbour' on average with whom one can exchange information), then all $N(\mathcal{A})$'s are small compared with the total number of traders $\mathcal{N}$. More precisely, the distribution of cluster sizes takes the following form in the limit where $1 - p = \epsilon \ll 1$:

$$P(N) \propto_{N \gg 1} \frac{1}{N^{5/2}} \exp(-\epsilon^2 N) \qquad N \ll \mathcal{N}. \tag{2.42}$$

When $p = 1$ (percolation threshold), the distribution becomes a pure power-law with an exponent $1 + \mu = \frac{5}{2}$, and the CLT tells us that in this case, the distribution of the price increments δx is precisely a pure symmetric Lévy distribution of index $\mu = \frac{3}{2}$ (assuming that $\varphi = \pm 1$ play identical roles, that is if there is no global bias pushing the price up or down). If $p < 1$, on the other hand, one finds that the Lévy distribution is truncated exponentially far in the tail. If $p > 1$, a finite fraction of the $\mathcal{N}$ traders have the same opinion: this leads to a crash.

This simple model is interesting but has one major drawback: one has to assume that the parameter p is smaller than one, but relatively close to one such that Eq. (2.42) is valid, and non-trivial statistics follows. One should thus explain why the value of p spontaneously stabilizes in the neighbourhood of the critical value $p = 1$. Certain models do actually have this property, of being close to or at a critical point without having to fine tune any of their parameters. These models are called 'self-organized critical' [Bak *et al.*]. In this spirit, let us mention a very recent model of Sethna *et al.* [Dahmen and Sethna], meant to describe the behaviour of magnets in a time dependent magnetic field. Transposed to the present problem, this model describes the collective behaviour of a set of traders exchanging information, but having all different *a priori* opinions. One trader can however change his mind and take the opinion of his neighbours if the coupling is strong, or if the strength of his *a priori* opinion is weak. All these traders feel an external 'field', which represents for example a long-term expectation of economy growth or recession, leading to an increased average pessimism or optimism. For a large range of parameters, one finds that the buy orders (or the sell orders) organize as avalanches of various sizes, distributed as a power-law with an exponential cut-off, with $\mu = \frac{5}{4} = 1.25$. If the anticipation of the traders are too similar, or if the coupling between agents is too strong (strong mimetism), the model again leads to a crash-like behaviour.

2.9 A simple model with volatility correlations and tails (*)

In this section, we show that a very simple feedback model where past high values of the volatility influence the present market activity does lead to tails in the probability distribution and, by construction, to volatility correlations. The present model is close in spirit to the ARCH models which have been much discussed in this context. The idea is to write:

$$x_{k+1} = x_k + \sigma_k \xi_k, \tag{2.43}$$

where ξ_k is a random variable of unit variance, and to postulate that the present day volatility σ_k depends on how the market feels the past market volatility. If the past price variations happened to be high, the market interprets this as a reason to be more nervous and increases its activity, thereby increasing σ_k. One could therefore consider, as a toy-model:[25]

$$\sigma_{k+1} - \sigma_0 = (1 - \epsilon)(\sigma_k - \sigma_0) + \lambda\epsilon|\sigma_k \xi_k|, \tag{2.44}$$

which means that the market takes as an indicator of the past day activity the absolute value of the close to close price difference $x_{k+1} - x_k$. Now, writing

$$|\sigma_k \xi_k| = \langle|\sigma\xi|\rangle + \sigma\tilde{\xi}, \tag{2.45}$$

and going to a continuous-time formulation, one finds that the volatility probability distribution $P(\sigma, t)$ obeys the following 'Fokker–Planck' equation:

$$\frac{\partial P(\sigma, t)}{\partial t} = \epsilon \frac{\partial(\sigma - \tilde{\sigma}_0)P(\sigma, t)}{\partial \sigma} + c^2\epsilon^2 \frac{\partial^2 \sigma^2 P(\sigma, t)}{\partial \sigma^2}, \tag{2.46}$$

where $\tilde{\sigma}_0 = \sigma_0 - \lambda\epsilon\langle|\sigma\xi|\rangle$, and where c^2 is the variance of the noise $\lambda\tilde{\xi}$. The equilibrium solution of this equation, $P_e(\sigma)$, is obtained by setting the left-hand

[25] In the simplest ARCH model, the following equation is rather written in terms of the variance, and second term of the right-hand side is taken to be equal to: $\epsilon(\sigma_k \xi_k)^2$.

side to zero. One finds:

$$P_e(\sigma) = \frac{\exp(-\tilde{\sigma}_0/\sigma)}{\sigma^{1+\mu}}, \tag{2.47}$$

with $\mu = 1 + (c^2\epsilon)^{-1} > 1$. Now, for a large class of distributions for the random noise ξ_k, for example Gaussian, it is easy to show, using a saddle-point calculation, that the tails of the distribution of δx are power-laws, with the same exponent μ. Interestingly, a short-memory market, corresponding to $\epsilon \simeq 1$, has much wilder tails than a long-memory market: in the limit $\epsilon \to 0$, one indeed has $\mu \to \infty$. In other words, over-reactions is a potential cause for power-law tails.

2.10 Conclusion

The above statistical analysis reveals very important differences between the simple model usually adopted to describe price fluctuations, namely the geometric (continuous-time) Brownian motion and the rather involved statistics of real price changes. The geometric Brownian motion description is at the heart of most theoretical work in mathematical finance, and can be summarized as follows:

- One assumes that the relative returns (rather than the absolute price increments) are independent random variables.
- One assumes that the elementary time scale τ *tends to zero*; in other words that the price process is a continuous-time process. It is clear that in this limit, the number of independent price changes in an interval of time T is $N = T/\tau \to \infty$. One is thus in the limit where the CLT applies whatever the time scale T.

If the variance of the returns is finite, then according to the CLT, the only possibility is that price changes obey a log-normal distribution. The process is also *scale invariant*, that is that its statistical properties do not depend on the chosen time scale (up to a multiplicative factor – see Section 1.5.3).[26]

The main difference between this model and real data is not only that the tails of the distributions are very poorly described by a Gaussian law, but also that several important time scales appear in the analysis of price changes:

- A 'microscopic' time scale τ below which price changes are correlated. This time scale is of the order of several minutes even on very liquid markets.
- A time scale $T^* = N^*\tau$, which corresponds to the time where non-Gaussian effects begin to smear out, beyond which the CLT begins to operate. This time scale T^* depends much on the initial kurtosis on scale τ. As a first approximation, one has: $T^* = \kappa_1\tau$, which is equal to several days, even on very liquid markets.
- A time scale corresponding to the correlation time of the volatility fluctuations, which is of the order of 10 days to a month or even longer.
- And finally a time scale T_σ governing the crossover from an additive model, where absolute price changes are the relevant random variables, to a multiplicative model, where relative returns become relevant. This time scale is also of the order of months.

[26] This scale invariance is more general than the Gaussian model discussed here, and is the basic assumption underlying all 'fractal' descriptions of financial markets. These descriptions fail to capture the existence of several important time scales that we discuss here.

It is clear that the existence of all these time scales is extremely important to take into account in a faithful representation of price changes, and play a crucial role both in the pricing of derivative products, and in risk control. Different assets differ in the value of their kurtosis, and in the value of these different time scales. For this reason, a description where the volatility is the only parameter (as is the case for Gaussian models) are bound to miss a great deal of the reality.

2.11 References

Scaling and Fractals in Financial Markets

B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982.
B. B. Mandelbrot, *Fractals and Scaling in Finance*, Springer, New York, 1997.
J. Lévy-Véhel, Ch. Walter, *Les marchés fractals : efficience, ruptures et tendances sur les marchés financiers*, P.U.F., to appear.
U. A. Muller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwartz, C. Morgenegg, Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis, *Journal of Banking and Finance*, **14**, 1189 (1990).
Z. Ding, C. W. J. Granger, R. F. Engle, A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, **1**, 83 (1993).
D. M. Guillaume *et al.*, From the birds eye to the microscope, *Finance and Stochastics*, **1**, 2 (1997).
S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Turbulent cascades in foreign exchange markets, *Nature*, **381**, 767 (1996).
A. Arnéodo, J. F. Muzy, D. Somette, Causal cascade in the stock market from the 'infrared' to the 'ultraviolet', *European Journal of Physics*, **B 2**, 277 (1998).
R. Mantegna, H. E. Stanley, *An Introduction to Econophysics*, Cambridge University Press, Cambridge, 1999.
I. Kondor, J. Kertecz (eds), *Econophysics, an Emerging Science*, Kluwer, Dordrecht, 1999.

Other descriptions (ARCH, etc.)

R. Engle, ARCH with estimates of the variance of the United Kingdom inflation, *Econometrica*, **50**, 987 (1982).
T. Bollerslev, Generalized ARCH, *Journal of Econometrics*, **31**, 307 (1986).
C. Gourieroux, A. Montfort, *Statistics and Econometric Models*, Cambridge University Press, Cambridge, 1996.
E. Eberlein, U. Keller, Option pricing with hyperbolic distributions, *Bernoulli*, **1**, 281 (1995).

The interest rate curve

L. Hughston (ed.), *Vasicek and Beyond*, Risk Publications, London, 1996.
J. C. Hull, *Options, Futures and Other Derivatives*, Prentice-Hall, Upper Saddle River, NJ, 1997.

Percolation, collective models and self organized criticality

D. Stauffer, *Introduction to Percolation*, Taylor & Francis, London, 1985.
A. Orléan, Bayesian interactions and collective dynamics of opinion, *Journal of Economic Behaviour and Organization*, **28**, 257 (1995).
P. Bak, *How Nature Works: The Science of Self-Organized Criticality*, Copernicus, Springer, New York, 1996.
K. Dahmen, J. P. Sethna, Hysteresis, avalanches, and disorder induced critical scaling, *Physical Review B*, **53**, 14 872 (1996).

Other recent market models

H. Takayasu, H. Miura, T. Hirabayashi, K. Hamada, Statistical properties of deterministic threshold elements: the case of market prices, *Physica*, **A184**, 127–134 (1992).
M. Lévy, H. Lévy, S. Solomon, Microscopic simulation of the stock market, *Journal de Physique*, **I5**, 1087 (1995).
D. Challet, Y. C. Zhang, Emergence of cooperation and organization in an evolutionary game, *Physica*, **A246**, 407 (1997).
P. Bak, M. Paczuski, M. Shubik, Price variations in a stock market with many agents, *Physica*, **A246**, 430 (1997).
J.-P. Bouchaud, R. Cont, A Langevin approach to stock market fluctuations and crashes, *European Journal of Physics*, **B 6**, 543 (1998).
T. Lux, M. Marchesi, Scaling and criticality in a stochastic multiagent model, *Nature*, **397**, 498 (1999).

3

Extreme risks and optimal portfolios

Il n'est plus grande folie que de placer son salut dans l'incertitude.[1]
(Madame de Sévigné, *Lettres*.)

3.1 Risk measurement and diversification

Measuring and controlling risks is now one of the major concern across all modern human activities. The financial markets, which act as highly sensitive economical and political thermometers, are no exception. One of their rôles is actually to allow the different actors in the economic world to *trade* their risks, to which a price must therefore be given.

The very essence of the financial markets is to fix thousands of prices all day long, thereby generating enormous quantities of data that can be analysed statistically. An objective measure of risk therefore appears to be easier to achieve in finance than in most other human activities, where the definition of risk is vaguer, and the available data often very poor. Even if a purely statistical approach to financial risks is itself a dangerous scientists' dream (see e.g. Fig. 1.1), it is fair to say that this approach has not been fully exploited until the very recent years, and that many improvements can be expected in the future, in particular concerning the control of extreme risks. The aim of this chapter is to introduce some classical ideas on financial risks, to illustrate their weaknesses, and to propose several theoretical ideas devised to handle more adequately the 'rare events' where the true financial risk resides.

3.1.1 Risk and volatility

Financial risk has been traditionally associated with the statistical uncertainty on the final outcome. Its traditional measure is the RMS, or, in financial terms, the

[1] Nothing is more foolish than betting on uncertainty for salvation.

volatility. We will note by $R(T)$ the logarithmic return on the time interval T, defined by:

$$R(T) = \log\left[\frac{x(T)}{x_0}\right], \tag{3.1}$$

where $x(T)$ is the price of the asset X at time T, knowing that it is equal to x_0 today ($t = 0$). When $|x(T) - x_0| \ll x_0$, this definition is equivalent to $R(T) = x(T)/x_0 - 1$.

If $P(x, T|x_0, 0)\,\mathrm{d}x$ is the conditional probability of finding $x(T) = x$ within $\mathrm{d}x$, the volatility σ of the investment is the standard deviation of $R(T)$, defined by:

$$\sigma^2 = \frac{1}{T}\left[\int P(x, T|x_0, 0)R^2(T)\,\mathrm{d}x - \left(\int P(x, T|x_0, 0)R(T)\,\mathrm{d}x\right)^2\right]. \tag{3.2}$$

The volatility is in general chosen as an adequate measure of risk associated to a given investment. We notice however that this definition includes in a symmetrical way both abnormal gains and abnormal losses. This fact is *a priori* curious. The theoretical foundations behind this particular definition of risk are numerous:

- First, operational; the computations involving the variance are relatively simple and can be generalized easily to multi-asset portfolios.
- Second, the Central Limit Theorem (CLT) presented in Chapter 1 seems to provide a general and solid justification: by decomposing the motion from x_0 to $x(T)$ in $N = T/\tau$ increments, one can write:

$$x(T) = x_0 + \sum_{k=0}^{N-1} \delta x_k \quad \text{with} \quad \delta x_k = x_k \eta_k, \tag{3.3}$$

where $x_k = x(t = k\tau)$ and η_k is by definition the instantaneous return. Therefore, we have:

$$R(T) = \log\left[\frac{x(T)}{x_0}\right] = \sum_{k=0}^{N-1} \log(1 + \eta_k). \tag{3.4}$$

In the classical approach one assumes that the returns η_k are independent variables. From the CLT we learn that *in the limit where* $N \to \infty$, $R(T)$ becomes a Gaussian random variable centred on a given average return $\tilde{m}T$, with $\tilde{m} = \langle\log(1 + \eta_k)\rangle/\tau$, and whose standard deviation is given by $\sigma\sqrt{T}$.[2] Therefore, in this limit, the entire probability distribution of $R(T)$ is parameterized by two quantities only, $\tilde{m}$ and σ: any reasonable measure of risk must therefore be based on σ.

[2] To second order in $\eta_k \ll 1$, we find: $\sigma^2 = \frac{1}{\tau}\langle\eta^2\rangle$ and $\tilde{m} = \frac{1}{\tau}\langle\eta\rangle - \frac{1}{2}\sigma^2$.

However, as we have discussed at length in Chapter 1, this is not true for finite N (which corresponds to the financial reality: there are only roughly $N \simeq 320$ half-hour intervals in a working month), especially in the 'tails' of the distribution, corresponding precisely to the extreme risks. We will discuss this point in detail below.

One can give to σ the following intuitive meaning: after a long enough time T, the price of asset X is given by:

$$x(T) = x_0 \exp[\tilde{m}T + \sigma\sqrt{T}\xi], \tag{3.5}$$

where ξ is a Gaussian random variable with zero mean and unit variance. The quantity $\sigma\sqrt{T}$ gives us the order of magnitude of the deviation from the expected return. By comparing the two terms in the exponential, one finds that when $T \gg \hat{T} \equiv \sigma^2/\tilde{m}^2$, the expected return becomes more important than the fluctuations, which means that the probability that $x(T)$ is smaller than x_0 (and that the actual rate of return over that period is negative) becomes small. The 'security horizon' $\hat{T}$ increases with σ. For a typical individual stock, one has $\tilde{m} = 10\%$ per year and $\sigma = 20\%$ per year, which leads to a $\hat{T}$ as long as 4 years!

The quality of an investment is often measured by its 'Sharpe ratio' $\mathcal{S}$, that is, the 'signal-to-noise' ratio of the mean return $\tilde{m}T$ to the fluctuations $\sigma\sqrt{T}$:[3]

$$\mathcal{S} = \frac{\tilde{m}\sqrt{T}}{\sigma} \equiv \sqrt{\frac{T}{\hat{T}}}. \tag{3.6}$$

The Sharpe ratio increases with the investment horizon and is equal to 1 precisely when $T = \hat{T}$. Practitioners usually define the Sharpe ratio for a 1-year horizon.

Note that the most probable value of $x(T)$, as given by Eq. (3.5), is equal to $x_0\exp(\tilde{m}T)$, whereas the mean value of $x(T)$ is higher: assuming that ξ is Gaussian, one finds: $x_0\exp[(\tilde{m}T + \sigma^2 T/2)]$. This difference is due to the fact that the *returns* η_k, rather than the absolute increments δx_k, are iid random variables. However, if T is short (say up to a few months), the difference between the two descriptions is hard to detect. As explained in Section 2.2.1, a purely additive description is actually more adequate at short times. In other words, we shall often in the following write $x(T)$ as:

$$x(T) = x_0\exp[\tilde{m}T + \sigma\sqrt{T}\xi] \simeq x_0 + mT + \sqrt{DT}\xi, \tag{3.7}$$

where we have introduced the following notations: $m \equiv \tilde{m}x_0$, $D \equiv \sigma^2 x_0^2$, which we shall use throughout the following. The non-Gaussian nature of the random variable ξ is therefore the most important factor determining the probability for extreme risks.

[3] It is customary to subtract from the mean return $\tilde{m}$ the risk-free rate in the definition of the Sharpe ratio.

3.1.2 Risk of loss and 'Value at Risk' (VaR)

The fact that financial risks are often described using the volatility is actually intimately related to the idea that the distribution of price changes is Gaussian. In the extreme case of 'Lévy fluctuations', for which the variance is infinite, this definition of risk would obviously be meaningless. Even in a less risky world, this measure of risk has three major drawbacks:

- The financial risk is obviously associated to losses and not to profits. A definition of risk where both events play symmetrical roles is thus not in conformity with the intuitive notion of risk, as perceived by professionals.
- As discussed at length in Chapter 1, a Gaussian model for the price fluctuations is *never justified* for the extreme events, since the CLT only applies in the *centre* of the distributions. Now, it is precisely these extreme risks that are of most concern for all financial houses, and thus those which need to be controlled in priority. In recent years, international regulators have tried to impose some rules to limit the exposure of banks to these extreme risks.
- The presence of extreme events in the financial time series can actually lead to a very bad empirical determination of the variance: its value can be substantially changed by a few 'big days'. A bold solution to this problem is simply to remove the contribution of these so-called aberrant events! This rather absurd solution is actually quite commonly used.

Both from a fundamental point of view, and for a better control of financial risks, another definition of risk is thus needed. An interesting notion that we shall develop now is the *probability of extreme losses*, or, equivalently, the 'value-at-risk' (VaR).

The probability to lose an amount $-\delta x$ larger than a certain threshold Λ on a given time horizon τ is defined as:

$$\mathcal{P}[\delta x < -\Lambda] = \mathcal{P}_<[-\Lambda] = \int_{-\infty}^{-\Lambda} P_\tau(\delta x)\,\mathrm{d}\delta x, \tag{3.8}$$

where $P_\tau(\delta x)$ is the probability density for a price change on the time scale τ. One can alternatively define the risk as a level of loss (the 'VaR') Λ_{VaR} corresponding to a certain probability of loss $\mathcal{P}_{\mathrm{VaR}}$ over the time interval τ (for example, $\mathcal{P}_{\mathrm{VaR}} = 1\%$):

$$\int_{-\infty}^{-\Lambda_{\mathrm{VaR}}} P_\tau(\delta x)\,\mathrm{d}\delta x = \mathcal{P}_{\mathrm{VaR}}. \tag{3.9}$$

This definition means that a loss greater than Λ_{VaR} over a time interval of $\tau = 1$ day (for example) happens only every 100 days on average for $\mathcal{P}_{\mathrm{VaR}} = 1\%$. Let us note that this definition does not take into account the fact that losses can accumulate on consecutive time intervals τ, leading to an overall loss which might substantially

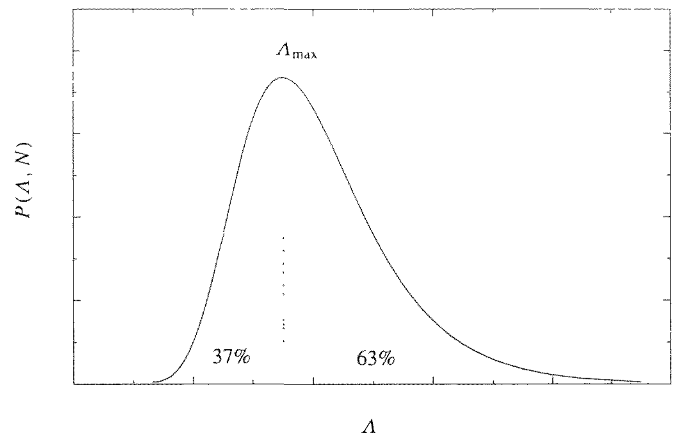

Fig. 3.1. Extreme value distribution (the so-called Gumbel distribution) $P(\Lambda, N)$ when $P_\tau(\delta x)$ decreases faster than any power-law. The most probable value, $\Lambda_{\max}$, has a probability equal to 0.63 to be exceeded.

exceed Λ_{VaR}. Similarly, this definition does not take into account the value of the maximal loss 'inside' the period τ. In other words, only the closing price over the period $[k\tau, (k+1)\tau]$ is considered, and not the lowest point reached during this time interval: we shall come back on these temporal aspects of risk in Section 3.1.3.

More precisely, one can discuss the probability distribution $P(\Lambda, N)$ for the worst daily loss Λ (we choose $\tau = 1$ day to be specific) on a temporal horizon $T_{\mathrm{VaR}} = N\tau = \tau/\mathcal{P}_{\mathrm{VaR}}$. Using the results of Section 1.4, one has:

$$P(\Lambda, N) = N[\mathcal{P}_>(-\Lambda)]^{N-1} P_\tau(-\Lambda). \tag{3.10}$$

For N large, this distribution takes a universal shape that only depends on the asymptotic behaviour of $P_\tau(\delta x)$ for $\delta x \to -\infty$. In the important case for practical applications where $P_\tau(\delta x)$ decays faster than any power-law, one finds that $P(\Lambda, N)$ is given by Eq. (1.40), which is represented in Figure 3.1. This distribution reaches a maximum precisely for $\Lambda = \Lambda_{\mathrm{VaR}}$, defined by Eq. (3.9). The intuitive meaning of Λ_{VaR} is thus the value of the most probable worst day over a time interval T_{VaR}. Note that the probability for Λ to be even worse ($\Lambda > \Lambda_{\mathrm{VaR}}$) is equal to 63% (Fig. 3.1). One could define Λ_{VaR} in such a way that this probability is smaller, by requiring a higher confidence level, for example 95%. This would mean that on a given horizon (for example 100 days), the probability that the worst day is found to be beyond Λ_{VaR} is equal to 5%. This Λ_{VaR} then corresponds to the most probable worst day on a time period equal to $100/0.05 = 2000$ days.

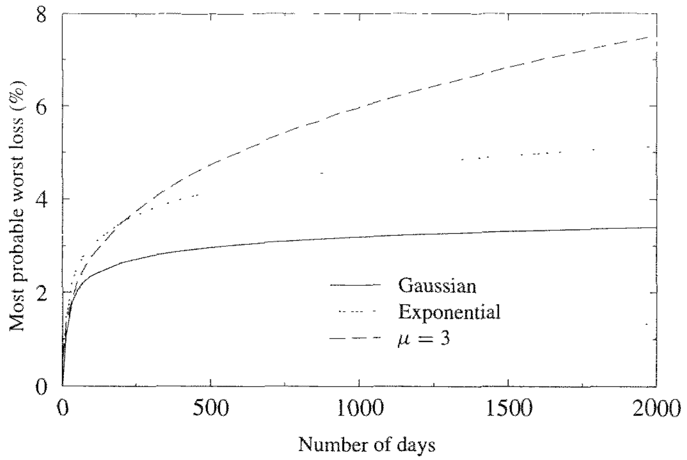

Fig. 3.2. Growth of Λ_{max} as a function of the number of days N, for an asset of daily volatility equal to 1%, with different distribution of price increments: Gaussian, (symmetric) exponential, or power-law with $\mu = 3$ (cf. Eq. (1.83)). Note that for intermediate N, the exponential distribution leads to a larger VaR than the power-law; this relation is however inverted for $N \to \infty$.

In the Gaussian case, the VaR is directly related to the volatility σ. Indeed, for a Gaussian distribution of RMS equal to $\sigma_1 x_0 = \sigma x_0\sqrt{\tau}$, and of mean m_1, one finds that $\Lambda_{\rm VaR}$ is given by:

$$\mathcal{P}_{G<}\left(-\frac{\Lambda_{\rm VaR}+m_1}{\sigma_1 x_0}\right) = \mathcal{P}_{\rm VaR} \to \Lambda_{\rm VaR} = \sqrt{2}\sigma_1 x_0 \mathrm{erfc}^{-1}[2\mathcal{P}_{\rm VaR}] - m_1, \quad (3.11)$$

(cf. Eq. (1.68)). When m_1 is small, minimizing $\Lambda_{\rm VaR}$ is thus equivalent to minimizing σ. It is furthermore important to note the very slow growth of $\Lambda_{\rm VaR}$ as a function of the time horizon in a Gaussian framework. For example, for $T_{\rm VaR} = 250\tau$ (1 market year), corresponding to $\mathcal{P}_{\rm VaR} = 0.004$, one finds that $\Lambda_{\rm VaR} \simeq 2.65\sigma_1 x_0$. Typically, for $\tau = 1$ day, $\sigma_1 = 1\%$, and therefore, the most probable worst day on a market year is equal to -2.65%, and grows only to -3.35% over a 10-year horizon!

In the general case, however, there is no useful link between σ and $\Lambda_{\rm VaR}$. In some cases, decreasing one actually increases the other (see below, Section 3.2.3 and 3.4). Let us nevertheless mention the Chebyshev inequality, often invoked by the volatility fans, which states that if σ exists,

$$\Lambda_{\rm VaR}^2 \leq \frac{\sigma_1^2 x_0^2}{\mathcal{P}_{\rm VaR}}. \quad (3.12)$$

Table 3.1. J.P. Morgan international bond indices (expressed in French Francs), analysed over the period 1989–93, and worst day observed day in 1994. The predicted numbers correspond to the most probable worst day Λ_{max}. The amplitude of the worst day with 95% confidence level is easily obtained, in the case of exponential tails, by multiplying Λ_{max} by a factor 1.53. The last line corresponds to a portfolio made of the 11 bonds with equal weight. All numbers are in per cent

Country	Worst day Log-normal	Worst day TLD	Worst day Observed
Belgium	0.92	1.14	1.11
Canada	2.07	2.78	2.76
Denmark	0.92	1.08	1.64
France	0.59	0.74	1.24
Germany	0.60	0.79	1.44
Great Britain	1.59	2.08	2.08
Italy	1.31	2.60	4.18
Japan	0.65	0.82	1.08
Netherlands	0.57	0.70	1.10
Spain	1.22	1.72	1.98
United States	1.85	2.31	2.26
Portfolio	0.61	0.80	1.23

This inequality suggests that in general, the knowledge of σ is tantamount to that of $\Lambda_{\rm VaR}$. This is however completely wrong, as illustrated in Figure 3.2. We have represented Λ_{max} (= $\Lambda_{\rm VaR}$ with $\mathcal{P}_{\rm VaR} = 0.63$) as a function of the time horizon $T_{\rm VaR} = N\tau$ for three distributions $P_\tau(\delta x)$ which all have exactly the same variance, but decay as a Gaussian, as an exponential, or as a power-law with an exponent $\mu = 3$ (cf. Eq. (1.83)). Of course, the slower the decay of the distribution, the faster the growth of $\Lambda_{\rm VaR}$ when $T_{\rm VaR} \to \infty$.

Table 3.1 shows, in the case of international bond indices, a comparison between the prediction of the most probable worst day using a Gaussian model, or using the observed exponential character of the tail of the distribution, and the actual worst day observed the following year. It is clear that the Gaussian prediction is systematically over-optimistic. The exponential model leads to a number which is seven times out of 11 below the observed result, which is indeed the expected result (Fig. 3.1).

Note finally that the measure of risk as a loss probability keeps its meaning even if the variance is infinite, as for a Lévy process. Suppose indeed that $P_\tau(\delta x)$ decays

very slowly when δx is very large, as:

$$P_\tau(\delta x) \underset{\delta x \to -\infty}{\simeq} \frac{\mu A^\mu}{|\delta x|^{1+\mu}}, \tag{3.13}$$

with $\mu < 2$, such that $\langle \delta x^2 \rangle = \infty$. A^μ is the 'tail amplitude' of the distribution P_τ; A gives the order of magnitude of the probable values of δx. The calculation of Λ_{VaR} is immediate and leads to:

$$\Lambda_{\text{VaR}} = A\mathcal{P}_{\text{VaR}}^{-1/\mu}, \tag{3.14}$$

which shows that in order to minimize the VaR one should minimize A^μ, independently of the probability level $\mathcal{P}_{\text{VaR}}$.

Note however that in this case, the most probable loss level is not equal to Λ_{VaR} but to $(1 + 1/\mu)^{1/\mu}\Lambda_{\text{VaR}}$. The previous Gumbel case again corresponds formally to the limit $\mu \to \infty$.

As we have noted in Chapter 1, A^μ is actually the natural generalization of the variance in this case. Indeed, the Fourier transform $\hat{P}_\tau(z)$ of P_τ behaves, for small z, as $\exp(-b_\mu A^\mu |z|^\mu)$ for $\mu < 2$ (b_μ is a certain numerical factor), and simply as $\exp(-D\tau z^2/2)$ for $\mu > 2$, where $A^2 = D\tau$ is precisely the variance of P_τ.

3.1.3 Temporal aspects: drawdown and cumulated loss

Worst low

A first problem which arises is the following: we have defined $\mathcal{P}$ as the probability for the loss observed at the end of the period $[k\tau, (k+1)\tau]$ to be at least equal to Λ. However, in general, a worse loss still has been reached *within* this time interval. What is then the probability that the worst point reached within the interval (the 'low') is a least equal to Λ? The answer is easy for symmetrically distributed increments (we thus neglect the average return $m\tau \ll \Lambda$, which is justified for small enough time intervals): this probability is simply equal to $2\mathcal{P}$,

$$\mathcal{P}[X_{\text{lo}} - X_{\text{op}} < -\Lambda] = 2\mathcal{P}[X_{\text{cl}} - X_{\text{op}} < -\Lambda]. \tag{3.15}$$

where X_{op} is the value at the beginning of the interval (open), X_{cl} at the end (close) and X_{lo}, the lowest value in the interval (low). The reason for this is that for each trajectory just reaching $-\Lambda$ between $k\tau$ and $(k+1)\tau$, followed by a path which ends up above $-\Lambda$ at the end of the period, there is a mirror path with precisely the same weight which reaches a 'close' value beyond $-\Lambda$.[4] This factor of 2 in cumulative probability is illustrated in Figure 3.3. Therefore, if one wants to take into account the possibility of further loss within the time interval of interest in

[4] In fact, this argument – which dates back to Bachelier himself (1900)! – assumes that the moment where the trajectory reaches the point $-\Lambda$ can be precisely identified. This is not the case for a discontinuous process, for which the doubling of $\mathcal{P}$ is only an approximate result.

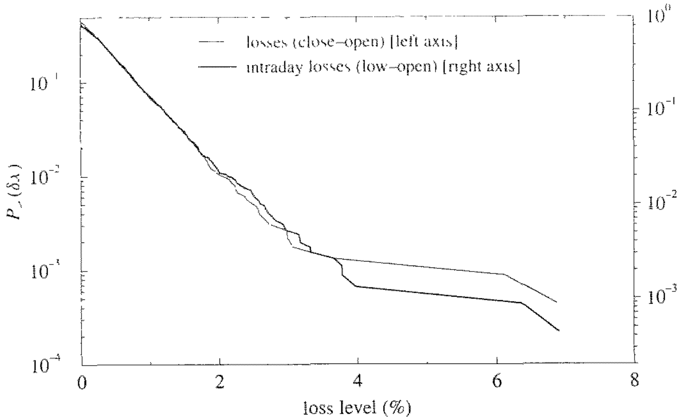

Fig. 3.3. Cumulative probability distribution of losses (Rank histogram) for the S&P 500 for the period 1989–98. Shown is the daily loss $(X_{\text{cl}} - X_{\text{op}})/X_{\text{op}}$ (thin line, left axis) and the intraday loss $(X_{\text{lo}} - X_{\text{op}})/X_{\text{op}}$ (thick line, right axis). Note that the right axis is shifted downwards by a factor of two with respect to the left one, so in theory the two lines should fall on top of one another.

Table 3.2. Average value of the absolute value of the open/close daily returns and maximum daily range (high–low) over open for S&P 500, DEM/$ and Bund. Note that the ratio of these two quantities is indeed close to 2. Data from 1989 to 1998

	$A = \left\langle \frac{\lvert X_{\text{cl}} - X_{\text{op}} \rvert}{X_{\text{op}}} \right\rangle$	$B = \left\langle \frac{X_{\text{hi}} - X_{\text{lo}}}{X_{\text{op}}} \right\rangle$	B/A
S&P 500	0.472%	0.924%	1.96
DEM/$	0.392%	0.804%	2.05
Bund	0.250%	0.496%	1.98

the computation of the VaR, one should simply divide by a factor 2 the probability level $\mathcal{P}_{\text{VaR}}$ appearing in Eq. (3.9).

A simple consequence of this 'factor 2' rule is the following: the average value of the maximal excursion of the price during time τ, given by the high minus the low over that period, is equal to twice the average of the absolute value of the variation from the beginning to the end of the period ($\langle|\text{open–close}|\rangle$). This relation can be tested on real data; the corresponding empirical factor is reported in Table 3.2.

Cumulated losses

Another very important aspect of the problem is to understand how losses can accumulate over successive time periods. For example, a bad day can be followed

by several other bad days, leading to a large overall loss. One would thus like to estimate the most probable value of the worst week, month, etc. In other words, one would like to construct the graph of $\Lambda_{\mathrm{VaR}}(N\tau)$, for a fixed overall investment horizon T.

The answer is straightforward in the case where the price increments are independent random variables. When the elementary distribution P_τ is Gaussian, $P_{N\tau}$ is also Gaussian, with a variance multiplied by N. At the same time, the number of different intervals of size $N\tau$ for a fixed investment horizon T decreases by a factor N. For large enough T, one then finds:

$$\Lambda_{\mathrm{VaR}}(N\tau)|_T \simeq \sigma_1 x_0 \sqrt{2N \log\left(\frac{T}{\sqrt{2\pi}N\tau}\right)}, \tag{3.16}$$

where the notation $|_T$ means that the investment period is fixed.[5] The main effect is then the $\sqrt{N}$ increase of the volatility, up to a small logarithmic correction.

The case where $P_\tau(\delta x)$ decreases as a power-law has been discussed in Chapter 1: for any N, the far tail remains a power-law (that is progressively 'eaten up' by the Gaussian central part of the distribution if the exponent μ of the power-law is greater than 2, but keeps its integrity whenever $\mu < 2$). For finite N, the *largest moves* will always be described by the power-law tail. Its amplitude A^μ is simply multiplied by a factor N (cf. Section 1.5.2). Since the number of independent intervals is divided by the same factor N, one finds:[6]

$$\Lambda_{\mathrm{VaR}}(N\tau)|_T = A\left(\frac{NT}{N\tau}\right)^{\frac{1}{\mu}}, \tag{3.17}$$

independently of N. Note however that for $\mu > 2$, the above result is only valid if $\Lambda_{\mathrm{VaR}}(N\tau)$ is located outside of the Gaussian central part, the size of which growing as $\sigma\sqrt{N}$ (Fig. 3.4). In the opposite case, the Gaussian formula (3.16) should be used.

One can of course ask a slightly different question, by fixing not the investment horizon T but rather the probability of occurrence. This amounts to multiplying both the period over which the loss is measured and the investment horizon by the same factor N. In this case, one finds that:

$$\Lambda_{\mathrm{VaR}}(N\tau) = N^{\frac{1}{\mu}}\Lambda_{\mathrm{VaR}}(\tau). \tag{3.18}$$

The growth of the value-at-risk with N is thus faster for small μ, as expected. The case of Gaussian fluctuations corresponds to $\mu = 2$.

5 One can also take into account the average return, $m_1 = \langle\delta x\rangle$. In this case, one must subtract to $\Lambda_{\mathrm{VaR}}(N\tau)|_T$ the quantity $-m_1 N$ (the potential losses are indeed smaller if the average return is positive).

6 cf. previous footnote.

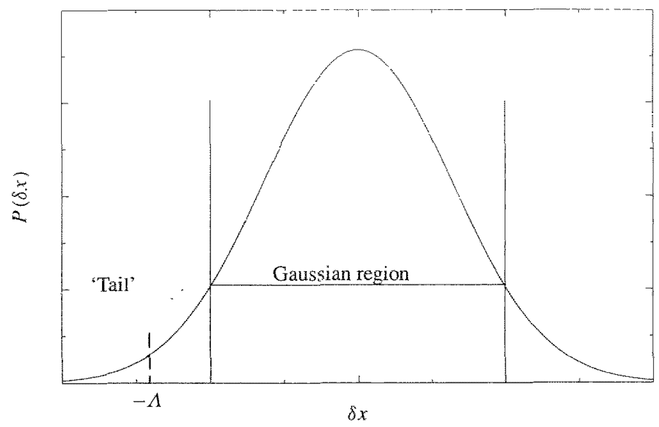

Fig. 3.4. Distribution of cumulated losses for finite N: the central region, of width $\sim\sigma N^{1/2}$, is Gaussian. The tails, however, remember the specific nature of the elementary distribution $P_\tau(\delta x)$, and are in general fatter than Gaussian. The point is then to determine whether the confidence level $\mathcal{P}_{\mathrm{VaR}}$ puts $\Lambda_{\mathrm{VaR}}(N\tau)$ within the Gaussian part or far in the tails.

Drawdowns

One can finally analyse the amplitude of a cumulated loss over a period that is not *a priori* limited. More precisely, the question is the following: knowing that the price of the asset today is equal to x_0, what is the probability that the lowest price ever reached in the future will be $x_{\min}$; and how long such a *drawdown* will last? This is a classical problem in probability theory [Feller, vol. II, p. 404]. Let us denote as $\Lambda_{\max} = x_0 - x_{\min}$ the maximum amplitude of the loss. If $P_\tau(\delta x)$ decays at least exponentially when $\delta x \to -\infty$, the result is that the tail of the distribution of $\Lambda_{\max}$ behaves as an exponential:

$$P(\Lambda_{\max}) \underset{\Lambda_{\max}\to\infty}{\propto} \exp\left(-\frac{\Lambda_{\max}}{\Lambda_0}\right), \tag{3.19}$$

where $\Lambda_0 > 0$ is the finite solution of the following equation:

$$\int \exp\left(-\frac{\delta x}{\Lambda_0}\right) P_\tau(\delta x)\,\mathrm{d}\delta x = 1. \tag{3.20}$$

Note that for this last equation to be meaningful, $P_\tau(\delta x)$ should decrease at least as $\exp(-|\delta x|/\Lambda_0)$ for large negative δx's. It is clear (see below) that if $P_\tau(\delta x)$ decreases as a power-law, say, the distribution of *cumulated losses* cannot decay exponentially, since it would then decay faster than that of individual losses!

It is interesting to show how this result can be obtained. Let us introduce the cumulative distribution $\mathcal{P}_>(\Lambda) = \int_\Lambda^\infty P(\Lambda_{\max})\,d\Lambda_{\max}$. *The first step of the walk, of size* δx *can either exceed* $-\Lambda$, *or stay above it. In the first case, the level* Λ *is immediately reached. In the second case, one recovers the very same problem as initially, but with* $-\Lambda$ *shifted to* $-\Lambda - \delta x$. *Therefore,* $\mathcal{P}_>(\Lambda)$ *obeys the following equation:*

$$\mathcal{P}_>(\Lambda) = \int_{-\infty}^{-\Lambda} P_\tau(\delta x)\,d\delta x + \int_{-\Lambda}^{+\infty} P_\tau(\delta x)\mathcal{P}_>(\Lambda + \delta x)\,d\delta x. \tag{3.21}$$

If one can neglect the first term in the right-hand side for large Λ *(which should be self-consistently checked), then, asymptotically,* $\mathcal{P}_>(\Lambda)$ *should obey the following equation:*

$$\mathcal{P}_>(\Lambda) = \int_{-\Lambda}^{+\infty} P_\tau(\delta x)\mathcal{P}_>(\Lambda + \delta x)\,d\delta x. \tag{3.22}$$

Inserting an exponential shape for $\mathcal{P}_>(\Lambda)$ *then leads to Eq. (3.20) for* Λ_0, *in the limit* $\Lambda \to \infty$. *This result is however only valid if* $P_\tau(\delta x)$ *decays sufficiently fast for large negative* δx*'s, such that Eq. (3.20) has a non-trivial solution.*

Let us study two simple cases:

- For Gaussian fluctuations, one has:

$$P_\tau(\delta x) = \frac{1}{\sqrt{2\pi D\tau}} \exp -\frac{(\delta x - m\tau)^2}{2D\tau}. \tag{3.23}$$

Equation (3.20) thus becomes:

$$-\frac{m}{\Lambda_0} + \frac{D}{2\Lambda_0^2} = 0 \to \Lambda_0 = \frac{D}{2m}. \tag{3.24}$$

Λ_0 gives the order of magnitude of the worst drawdown. One can understand the above result as follows: Λ_0 is the amplitude of the probable fluctuation over the characteristic time scale $\hat{T} = D/m^2$ introduced above. By definition, for times shorter than $\hat{T}$, the average return m is negligible. Therefore, one has: $\Lambda_0 \propto \sqrt{D\hat{T}} = D/m$.

If $m = 0$, the very idea of worst drawdown loses its meaning: if one waits a long enough time, the price can then reach arbitrarily low values. It is natural to introduce a quality factor Q that compares the average return $m_1 = m\tau$ to the amplitude of the worst drawdown Λ_0. One thus has $Q = m_1/\Lambda_0 = 2m^2\tau/D \equiv 2\tau/\hat{T}$. The larger the quality factor, the smaller the time needed to end a drawdown period.

- The case of exponential tails is important in practice (cf. Chapter 2). Choosing for simplicity $P_\tau(\delta x) = (2\alpha)^{-1}\exp(-\alpha|\delta x - m\tau|)$, the equation for Λ_0 is found to be:

$$\frac{\alpha^2}{\alpha^2 - \Lambda_0^{-2}} \exp\left(-\frac{m\tau}{\Lambda_0}\right) = 1. \tag{3.25}$$

In the limit where $m\tau\alpha \ll 1$ (i.e. that the potential losses over the time interval τ are much larger than the average return over the same period), one finds: $\Lambda_0 = 1/m\tau\alpha^2$.

- More generally, if the width of $P_\tau(\delta x)$ is much smaller than Λ_0, one can expand the exponential and find that the above equation $-(m/\Lambda_0) + (D/2\Lambda_0^2) = 0$ is still valid.

How long will these drawdowns last? The answer can only be statistical. The probability for the time of first return T to the initial investment level x_0, which one can take as the definition of a drawdown (although it could of course be immediately followed by a second drawdown), has, for large T, the following form:

$$P(T) \simeq \frac{\tau^{1/2}}{T^{3/2}} \exp\left(-\frac{T}{\hat{T}}\right). \tag{3.26}$$

The probability for a drawdown to last much longer than $\hat{T}$ is thus very small. In this sense, $\hat{T}$ appears as the characteristic drawdown time. Note that it is *not* equal to the *average* drawdown time, which is on the order of $\sqrt{\tau\hat{T}}$, and thus much smaller than $\hat{T}$. This is related to the fact that *short* drawdowns have a large probability: as τ decreases, a large number of drawdowns of order τ appear, thereby reducing their average size.

3.1.4 Diversification and utility – satisfaction thresholds

It is intuitively clear that one should not put all his eggs in the same basket. A *diversified* portfolio, composed of different assets with small mutual correlations, is less risky because the gains of some of the assets more or less compensate the loss of the others. Now, an investment with a small risk and small return must sometimes be preferred to a high yield, but very risky, investment.

The theoretical justification for the idea of diversification comes again from the CLT. A portfolio made up of M uncorrelated assets and of equal volatility, with weight $1/M$, has an overall volatility reduced by a factor $\sqrt{M}$. Correspondingly, the amplitude (and duration) of the worst drawdown is divided by a factor M (cf. Eq. (3.24)), which is obviously satisfying.

This qualitative idea stumbles over several difficulties. First of all, the fluctuations of financial assets are in general *strongly correlated*; this substantially decreases the possibility of true diversification, and requires a suitable theory to deal with these correlations and construct an 'optimal' portfolio: this is Markowitz's theory, to be detailed below. Furthermore, since price fluctuations can be strongly non-Gaussian, a volatility-based measure of risk might be unadapted: one should rather try to minimize the value-at-risk of the portfolio. It is therefore interesting to

look for an extension of the classical formalism, allowing one to devise *minimum VaR* portfolios. This will be presented in the next sections.

Now, one is immediately confronted with the problem of defining properly an 'optimal' portfolio. Usually, one invokes the rather abstract concept of 'utility functions', on which we shall briefly comment in this section, in particular to show that it does not naturally accommodate for the notion of value-at-risk.

We will call W_T the wealth of a given operator at time $t = T$. If one argues that the level of satisfaction of this operator is quantified by a certain function of W_T only,[7] which one usually calls the 'utility function' $U(W_T)$. This function is furthermore taken to be continuous and even twice differentiable. The postulated 'rational' behaviour for the operator is then to look for investments which maximize his expected utility, averaged over all possible histories of price changes:

$$\langle U(W_T)\rangle = \int P(W_T)U(W_T)\,\mathrm{d}W_T. \tag{3.27}$$

The utility function should be non-decreasing: a larger profit is clearly always more satisfying. One can furthermore consider the case where the distribution $P(W_T)$ is sharply peaked around its mean value $\langle W_T\rangle = W_0 + mT$. Performing a Taylor expansion of $\langle U(W_T)\rangle$ around $U(W_0 + mT)$ to second order, one deduces that the utility function must be such that:

$$\frac{\mathrm{d}^2 U}{\mathrm{d}W_T^2} < 0. \tag{3.28}$$

This property reflects the fact that for the same average return, a less risky investment should always be preferred.

A simple example of utility function compatible with the above constraints is the exponential function $U(W_T) = -\exp[-W_T/w_0]$. Note that w_0 has the dimensions of a *wealth*, thereby fixing a wealth scale in the problem. A natural candidate is the initial wealth of the operator, $w_0 \propto W_0$. If $P(W_T)$ is Gaussian, of mean $W_0 + mT$ and variance DT, one finds that the expected utility is given by:

$$\langle U\rangle = -\exp\left[-\frac{W_0}{w_0} - \frac{T}{w_0}\left(m - \frac{D}{2w_0}\right)\right]. \tag{3.29}$$

One could think of constructing a utility function with no intrinsic wealth scale by choosing a power-law: $U(W_T) = (W_T/w_0)^\alpha$ with $\alpha < 1$ to ensure the correct convexity. Indeed, in this case a change of w_0 can be reabsorbed in a change of scale

[7] But *not* of the whole 'history' of his wealth between $t = 0$ and T. One thus assumes that the operator is insensitive to what can happen between these two dates; this is not very realistic. One could however generalize the concept of utility function and deal with utility *functionals* $U(\{W(t)\}_{0\le t\le T})$.

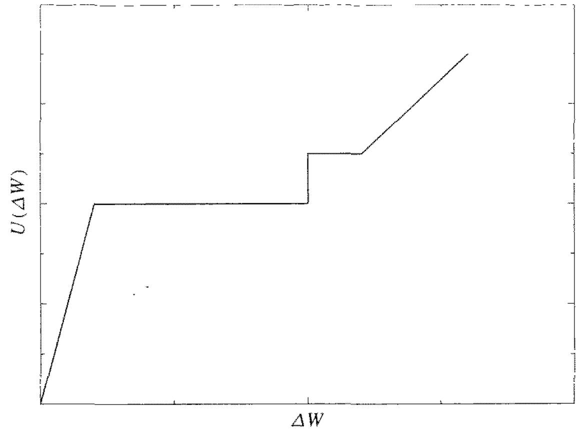

Fig. 3.5. Example of a 'utility function' with thresholds, where the utility function is non-continuous. These thresholds correspond to special values for the profit, or the loss, and are often of purely psychological origin.

of the utility function itself. However, this definition cannot allow for negative final wealths, and is thus problematic.

Despite the fact that these postulates sound reasonable, and despite the very large number of academic studies based on the concept of utility function, this axiomatic approach suffers from a certain number of fundamental flaws. For example, it is not clear that one could ever measure the utility function used by a given agent on the markets.[8] The theoretical results are thus:

- Either relatively weak, because independent of the special form of the utility function, and only based on its general properties.
- Or rather arbitrary, because based on a specific, but unjustified, form for $U(W_T)$.

On the other hand, the idea that the utility function is regular is probably not always realistic. The satisfaction of an operator is often governed by rather sharp *thresholds*, separated by regions of indifference (Fig. 3.5). For example, one can be in a situation where a specific project can only be achieved if the profit $\Delta W = W_T - W_0$ exceeds a certain amount. Symmetrically, the clients of a fund manager will take their money away as soon as the losses exceed a certain value: this is the strategy of 'stop-losses', which fix a level for acceptable losses, beyond which the position is closed. The existence of option markets (which allow one to limit

[8] Even the idea that an operator would really optimize his expected utility, and not take decisions partly based on 'non-rational' arguments, is far from being obvious. On this point, see: M. Marsili, Y. C. Zhang, Fluctuations around Nash equilibria in Game Theory, *Physica* **A245**, 181 (1997).

the potential losses below a certain level – see Chapter 4), or of items the price of which is \$99 rather than \$100, are concrete examples of the existence of these thresholds where 'satisfaction' changes abruptly. Therefore, the utility function U is not necessarily continuous. This remark is actually intimately related to the fact that the value-at-risk is often a better measure of risk than the variance. Let us indeed assume that the operator is 'happy' if $\Delta W > -\Lambda$ and 'unhappy' whenever $\Delta W < -\Lambda$. This corresponds formally to the following utility function:

$$U_\Lambda(\Delta W) = \begin{cases} U_1 & (\Delta W > -\Lambda) \\ U_2 & (\Delta W < -\Lambda) \end{cases}, \tag{3.30}$$

with $U_2 - U_1 < 0$.

The expected utility is then simply related to the loss probability:

$$\begin{aligned} \langle U_\Lambda \rangle &= U_1 + (U_2 - U_1) \int_{-\infty}^{-\Lambda} P(\Delta W)\, \mathrm{d}\Delta W \\ &= U_1 - |U_2 - U_1| \mathcal{P}. \end{aligned} \tag{3.31}$$

Therefore, optimizing the expected utility is in this case tantamount to minimizing the probability of losing more that Λ. Despite this rather appealing property, which certainly corresponds to the behaviour of some market operators, the function $U_\Lambda(\Delta W)$ does not satisfy the above criteria (continuity and negative curvature).

Confronted to the problem of choosing between risk (as measured by the variance) and return, another very natural strategy (for those not acquainted with utility functions) would be to compare the average return to the potential loss $\sqrt{DT}$. This can thus be thought of as defining a risk-corrected, 'pessimistic' estimate of the profit, as:

$$m_\lambda T = mT - \lambda\sqrt{DT}, \tag{3.32}$$

where λ is an arbitrary coefficient that measures the pessimism (or the risk aversion) of the operator. A rather natural procedure would then be to look for the optimal portfolio which maximizes the risk corrected return m_λ. However, this optimal portfolio cannot be obtained using the standard utility function formalism. For example, Eq. (3.29) shows that the object which should be maximized is *not* $mT - \lambda\sqrt{DT}$ but rather $mT - DT/2w_0$. This amounts to comparing the average profit to the *square* of the potential losses, divided by the reference wealth scale w_0, a quantity that depends *a priori* on the operator.[9] On the other hand, the quantity $mT - \lambda\sqrt{DT}$ is directly related (at least in a Gaussian world) to the value-at-risk Λ_{VaR}, cf. Eq. (3.16).

[9] This comparison is actually meaningful, since it corresponds to comparing the reference wealth w_0 to the order of magnitude of the worst drawdown D/m, cf. Eq. (3.24).

This can be expressed slightly differently: a reasonable objective could be to maximize the value of the 'probable gain' $\mathcal{G}_p$, such that the probability of earning more is equal to a certain probability p:[10]

$$\int_{\mathcal{G}_p}^{+\infty} P(\Delta W)\, \mathrm{d}\Delta W = p. \tag{3.33}$$

In the case where $P(\Delta W)$ is Gaussian, this amounts to maximizing $mT - \lambda\sqrt{DT}$, where λ is related to p in a simple manner. Now, one can show that it is impossible to construct a utility function such that, in the general case, different strategies can be ordered according to their probable gain $\mathcal{G}_p$. Therefore, the concepts of loss probability, value-at-risk or probable gain cannot be accommodated naturally within the framework of utility functions. Still, the idea that the quantity which is of most concern and that should be optimized is the value-at-risk sounds perfectly rational. This is at least the conceptual choice that we make in the present monograph.

3.1.5 Conclusion

Let us now recapitulate the main points of this section:

- The usual measure of risk through a Gaussian volatility is not always adapted to the real world. The tails of the distributions, where the large events lie, are very badly described by a Gaussian law: this leads to a systematic underestimation of the extreme risks. Sometimes, the measurement of the volatility on historical data is difficult, precisely because of the presence of these large fluctuations.
- The measure of risk through the probability of loss, or the value-at-risk, on the other hand, precisely focuses on the tails. Extreme events are considered as the true source of risk, whereas the small fluctuations contribute to the 'centre' of the distributions (and contribute to the volatility) can be seen as a background noise, inherent to the very activity of financial markets, but not relevant for risk assessment.
- From a theoretical point of view, this definition of risk (based on extreme events) does not easily fit into the classical 'utility function' framework. The minimization of a loss probability rather assumes that there exists well-defined thresholds (possibly different for each operator) where the 'utility function' is discontinuous.[11] The concept of 'value-at-risk', or probable gain, cannot be naturally dealt with by using utility functions.

[10] Maximizing $\mathcal{G}_p$ is thus equivalent to minimizing Λ_{VaR} such that $\mathcal{P}_{\mathrm{VaR}} = 1 - p$.
[11] It is possible that the presence of these thresholds actually plays an important role in the fact that the price fluctuations are strongly non-Gaussian.

3.2 Portfolios of uncorrelated assets

The aim of this section is to explain, in the very simple case where all assets that can be mixed in a portfolio are uncorrelated, how the trade-off between risk and return can be dealt with. (The case where some correlations between the asset fluctuations exist will be considered in the next section). On thus considers a set of M different risky assets X_i, $i = 1, \ldots, M$ and one risk-less asset X_0. The number of asset i in the portfolio is n_i, and its present value is x_i^0. If the total wealth to be invested in the portfolio is W, then the n_i's are constrained to be such that $\sum_{i=0}^M n_i x_i^0 = W$. We shall rather use the weight of asset i in the portfolio, defined as: $p_i = n_i x_i^0/W$, which therefore must be normalized to one: $\sum_{i=0}^M p_i = 1$. The p_i's can be negative (short positions). The value of the portfolio at time T is given by: $S = \sum_{i=0}^M n_i x_i(T) = W \sum_{i=0}^M p_i x_i(T)/x_i^0$. In the following, we will set the initial wealth W to 1, and redefine each asset i in such a way that all initial prices are equal to $x_i^0 = 1$. (Therefore, the average return m_i and variance D_i that we will consider below must be understood as relative, rather than absolute.)

One furthermore *assumes* that the average return m_i is known. This hypothesis is actually very strong, since it assumes for example that past returns can be used as estimators of future returns, i.e. that time series are to some extent stationary. However, this is very far from the truth: the life of a company (in particular high-tech ones) is very clearly non-stationary; a whole sector of activity can be booming or collapsing, depending upon global factors, not graspable within a purely statistical framework. Furthermore, the markets themselves evolve with time, and it is clear that some statistical parameters do depend on time, and have significantly shifted over the past 20 years. This means that the empirical determination of the average return is difficult: volatilities are such that at least several years are needed to obtain a reasonable signal-to-noise ratio, this time must indeed be large compared to the 'security time' $\hat{T}$. But as discussed above, several years is also the time scale over which the intrinsically non-stationary nature of the markets starts being important.

One should thus rather understand m_i as an 'expected' (or anticipated) future return, which includes some extra information (or intuition) available to the investor. These m_i's can therefore vary from one investor to the next. The relevant question is then to determine the composition of an optimal portfolio compatible with the information contained in the knowledge of the different m_i's.

The determination of the risk parameters is *a priori* subject to the same *caveat*. We have actually seen in Section 2.7 that the empirical determination of the correlation matrices contains a large amount of noise, which blur the true information. However, the statistical nature of the *fluctuations* seems to be more robust in time than the average returns. The analysis of past price changes distributions appears

to be, to a certain extent, predictive for future price fluctuations. It, however, sometimes happens that the correlations between two assets change abruptly.

3.2.1 Uncorrelated Gaussian assets

Let us suppose that the variation of the value of the ith asset X_i over the time interval T is Gaussian, centred around $m_i T$ and of variance $D_i T$. The portfolio $\mathbf{p} = \{p_0, p_1, \ldots, p_M\}$ as a whole also obeys Gaussian statistics (since the Gaussian is stable). The average return m_p of the portfolio is given by:

$$m_p = \sum_{i=0}^{M} p_i m_i = m_0 + \sum_{i=1}^{M} p_i (m_i - m_0), \tag{3.34}$$

where we have used the constraint $\sum_{i=0}^M p_i = 1$ to introduce the excess return $m_i - m_0$, as compared to the risk-free asset ($i = 0$). If the X_i's are all independent, the total variance of the portfolio is given by:

$$D_p = \sum_{i=1}^{M} p_i^2 D_i, \tag{3.35}$$

(since D_0 is zero by assumption). If one tries to minimize the variance without any constraint on the average return m_p, one obviously finds the trivial solution where all the weight is concentrated on the risk-free asset:

$$p_i = 0 \quad (i \neq 0); \quad p_0 = 1. \tag{3.36}$$

On the opposite, the maximization of the return without any risk constraint leads to a full concentration of the portfolio on the asset with the highest return.

More realistically, one can look for a tradeoff between risk and return, by imposing a certain average return m_p, and by looking for the less risky portfolio (for Gaussian assets, risk and variance are identical). This can be achieved by introducing a Lagrange multiplier in order to enforce the constraint on the average return:

$$\left.\frac{\partial(D_p - \zeta m_p)}{\partial p_i}\right|_{p_i = p_i^*} = 0 \quad (i \neq 0), \tag{3.37}$$

while the weight of the risk-free asset p_0^* is determined *via* the equation $\sum_i p_i^* = 1$. The value of ζ is ultimately fixed such that the average value of the return is precisely m_p. Therefore:

$$2 p_i^* D_i = \zeta (m_i - m_0), \tag{3.38}$$

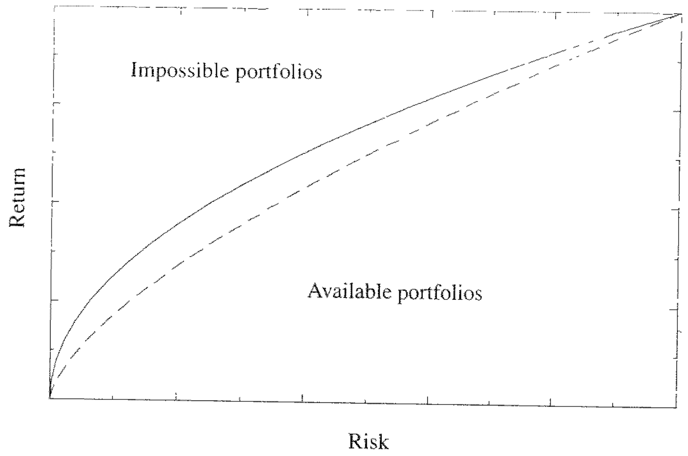

Fig. 3.6. 'Efficient frontier' in the return/risk plane m_p, D_p. In the absence of constraints, this line is a parabola (dark line). If some constraints are imposed (for example, that all the weights p_i should be positive), the boundary moves downwards (dotted line).

and the equation for ζ:

$$m_p - m_0 = \frac{\zeta}{2} \sum_{i=1}^{M} \frac{(m_i - m_0)^2}{D_i}. \tag{3.39}$$

The variance of this optimal portfolio is therefore given by:

$$D_p^* = \frac{\zeta^2}{4} \sum_{i=1}^{M} \frac{(m_i - m_0)^2}{D_i}. \tag{3.40}$$

The case $\zeta = 0$ corresponds to the risk-free portfolio $p_0^* = 1$. The set of all optimal portfolios is thus described by the parameter ζ, and define a *parabola* in the m_p, D_p plane (compare the last two equations below, and see Fig. 3.6). This line is called the 'efficient frontier'; all portfolios must lie above this line. Finally, the Lagrange multiplier ζ itself has a direct interpretation: Equation (3.24) tells us that the worst drawdown is of order $D_p^*/2m_p$, which is, using the above equations, equal to $\zeta/4$.

The case where the portfolio only contains risky assets (i.e. $p_0 \equiv 0$) can be treated in a similar fashion. One introduces a second Lagrange multiplier ζ' to deal with the normalization constraint $\sum_{i=1}^{M} p_i = 1$. Therefore, one finds:

$$p_i^* = \frac{\zeta m_i + \zeta'}{2D_i}. \tag{3.41}$$

The least risky portfolio corresponds to the one such that $\zeta = 0$ (no constraint on

the average return):

$$p_i^* = \frac{1}{ZD_i} \qquad Z = \sum_{j=1}^{M} \frac{1}{D_j}. \tag{3.42}$$

Its total variance is given by $D_p^* = 1/Z$. If all the D_i's are of the same order of magnitude, one has $Z \sim M/D$; therefore, one finds the result, expected from the CLT, that the variance of the portfolio is M times smaller than the variance of the individual assets.

In practice, one often adds extra constraints to the weights p_i^* in the form of linear inequalities, such as $p_i^* \geq 0$ (no short positions). The solution is then more involved, but is still unique. Geometrically, this amounts to looking for the restriction of a paraboloid to an hyperplane, which remains a paraboloid. The efficient border is then shifted downwards (Fig. 3.6). A much richer case is when the constraint is *non-linear*. For example, on futures markets, margin calls require that a certain amount of money is left as a deposit, whether the position is long ($p_i > 0$) or short ($p_i < 0$). One can then impose a *leverage constraint*, such that $\sum_{i=1}^{M} |p_i| = f$, where f is the fraction of wealth invested as a deposit. This constraint leads to a much more complex problem, similar to the one encountered in hard optimization problems, where an exponentially large (in M) number of quasi-degenerate solutions can be found.[12]

Effective asset number in a portfolio

It is useful to introduce an objective way to measure the diversification, or the asset concentration, in a given portfolio. Once such an indicator is available, one can actually use it as a constraint to construct portfolios with a minimum degree of diversification. Consider the quantity Y_2 defined as:

$$Y_2 = \sum_{i=1}^{M} (p_i^*)^2. \tag{3.43}$$

If a subset $M' \leq M$ of all p_i^* are equal to $1/M'$, while the others are zero, one finds $Y_2 = 1/M'$. More generally, Y_2 represents the average weight of an asset in the portfolio, since it is constructed as the average of p_i^* itself. It is thus natural to define the 'effective' number of assets in the portfolio as $M_{\text{eff}} = 1/Y_2$. In order to avoid an overconcentration of the portfolio on very few assets (a problem often encountered in practice), one can look for the optimal portfolio with a given value for Y_2. This amounts to introducing another Lagrange multiplier ζ'', associated to

[12] On this point, see S. Galluccio, J.-P. Bouchaud, M. Potters, Portfolio optimisation, spin-glasses and random matrix theory, *Physica*, **A259**, 449 (1998), and references therein.

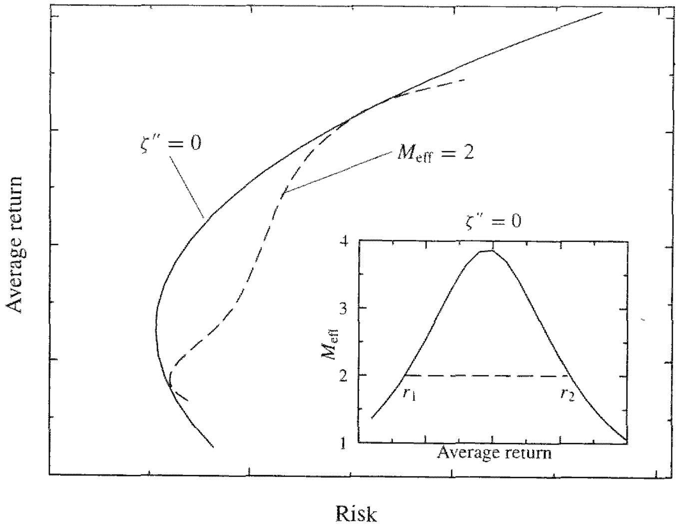

Fig. 3.7. Example of a standard efficient border $\zeta = 0$ (thick line) with four risky assets. If one imposes that the effective number of assets is equal to 2, one finds the sub-efficient border drawn in dotted line, which touches the efficient border at r_1, r_2. The inset shows the effective asset number of the unconstrained optimal portfolio ($\zeta'' = 0$) as a function of average return. The optimal portfolio satisfying $M_{\text{eff}} \geq 2$ is therefore given by the standard portfolio for returns between r_1 and r_2 and by the $M_{\text{eff}} = 2$ portfolios otherwise.

Y_2. The equation for p_i^* then becomes:

$$p_i^* = \frac{\zeta m_i + \zeta'}{2(D_i + \zeta'')}. \tag{3.44}$$

An example of the modified efficient border is given in Figure 3.7.

More generally, one could have considered the quantity Y_q defined as:

$$Y_q = \sum_{i=1}^{M} (p_i^*)^q, \tag{3.45}$$

and used it to define the effective number of assets via $Y_q = M_{\text{eff}}^{1-q}$. It is interesting to note that the quantity of missing information (or entropy) $\mathcal{I}$ associated to the very choice of the p_i^*'s is related to Y_q when $q \to 1$. Indeed, one has:

$$\mathcal{I} = -\sum_{i=1}^{M} p_i^* \log p_i^* \equiv -\left.\frac{\partial Y_q}{\partial q}\right|_{q=1}. \tag{3.46}$$

Approximating Y_q as a function of q by a straight line thus leads to $\mathcal{I} \simeq -Y_2$.

3.2.2 *Uncorrelated 'power-law' assets*

As we have already underlined, the tails of the distributions are often non-Gaussian. In this case, the minimization of the variance is not necessarily equivalent to an optimal control of the large fluctuations. The case where these distribution tails are power-laws is interesting because one can then explicitly solve the problem of the minimization of the value-at-risk of the full portfolio. Let us thus assume that the fluctuations of each asset X_i are described, in the region of large losses, by a probability density that decays as a power-law:

$$P_T(\delta x_i) \underset{\delta x_i \to -\infty}{\simeq} \frac{\mu A_i^\mu}{|\delta x_i|^{1+\mu}}, \tag{3.47}$$

with an arbitrary exponent μ, restricted however to be larger than 1, such that the average return is well defined. (The distinction between the cases $\mu < 2$, for which the variance diverges, and $\mu > 2$ will be considered below). The coefficient A_i provides an order of magnitude for the extreme losses associated with the asset i (cf. Eq. (3.14)).

As we have mentioned in Section 1.5.2, the power-law tails are interesting because they are stable upon addition: the tail amplitudes A_i^μ (that generalize the variance) simply add to describe the far-tail of the distribution of the sum. Using the results of Appendix C, one can show that if the asset X_i is characterized by a tail amplitude A_i^μ, the quantity $p_i X_i$ has a tail amplitude equal to $p_i^\mu A_i^\mu$. The tail amplitude of the global portfolio **p** is thus given by:

$$A_p^\mu = \sum_{i=1}^{M} p_i^\mu A_i^\mu, \tag{3.48}$$

and the probability that the loss exceeds a certain level Λ is given by $\mathcal{P} = A_p^\mu / \Lambda^\mu$. Hence, independently of the chosen loss level Λ, the minimization of the loss probability $\mathcal{P}$ requires the minimization of the tail amplitude A_p^μ; the optimal portfolio is therefore independent of Λ. (This is *not* true in general: see Section 3.2.4.) The minimization of A_p^μ for a fixed average return m_p leads to the following equations (valid if $\mu > 1$):

$$\mu p_i^{*\mu-1} A_i^\mu = \zeta(m_i - m_0), \tag{3.49}$$

with an equation to fix ζ:

$$\left(\frac{\zeta}{\mu}\right)^{\frac{1}{\mu-1}} \sum_{i=1}^{M} \frac{(m_i - m_0)^{\frac{\mu}{\mu-1}}}{A_i^{\frac{\mu}{\mu-1}}} = m_p - m_0. \tag{3.50}$$

The optimal loss probability is then given by:

$$\mathcal{P}^* = \frac{1}{A^\mu}\left(\frac{\zeta}{\mu}\right)^{\frac{\mu}{\mu-1}} \sum_{i=1}^{M} \frac{(m_i - m_0)^{\frac{\mu}{\mu-1}}}{A_i^{\frac{\mu}{\mu-1}}}. \tag{3.51}$$

Therefore, the concept of 'efficient border' is still valid in this case: in the plane return/probability of loss, it is similar to the dotted line of Figure 3.6. Eliminating ζ from the above two equations, one finds that the shape of this line is given by $\mathcal{P}^* \propto (m_p - m_0)^\mu$. The parabola is recovered in the limit $\mu = 2$.

In the case where the risk-free asset cannot be included in the portfolio, the optimal portfolio which minimizes extreme risks with no constraint on the average return is given by:

$$p_i^* = \frac{1}{Z A_i^{\frac{\mu}{\mu-1}}} \qquad Z = \sum_{j=1}^{M} A_i^{-\frac{\mu}{\mu-1}}, \tag{3.52}$$

and the corresponding loss probability is equal to:

$$\mathcal{P}^* = \frac{1}{A^\mu} Z^{1-\mu}. \tag{3.53}$$

If all assets have comparable tail amplitudes $A_i \sim A$, one finds that $Z \sim M A^{-\mu/(\mu-1)}$. Therefore, the probability of large losses for the optimal portfolio is a factor $M^{\mu-1}$ smaller than the individual probability of loss.

Note again that this result is only valid if $\mu > 1$. If $\mu < 1$, one finds that the risk increases with the number of assets M. In this case, when the number of assets is increased, the probability of an unfavourable event also increases – indeed, for $\mu < 1$ this largest event is so large that it dominates over all the others. The minimization of risk in this case leads to $p_{i_{\min}} = 1$, where $i_{\min}$ is the least risky asset, in the sense that $A_{i_{\min}}^\mu = \min\{A_i^\mu\}$.

One should now distinguish the cases $\mu < 2$ and $\mu > 2$. Despite the fact that the asymptotic power-law behaviour is stable under addition for all values of μ, the tail is progressively 'eaten up' by the centre of the distribution for $\mu > 2$, since the CLT applies. Only when $\mu < 2$ does this tail remain untouched. We thus again recover the arguments of Section 1.6.4, already encountered when we discussed the time dependence of the VaR. One should therefore distinguish two cases: if D_p is the variance of the portfolio **p** (which is finite if $\mu > 2$), the distribution of the changes δS of the value of the portfolio **p** is approximately Gaussian if $|\delta S| \le \sqrt{D_p T \log(M)}$, and becomes a power-law with a tail amplitude given by A_p^μ beyond this point. The question is thus whether the loss level Λ that one wishes to control is smaller or larger than this crossover value:

- If $\Lambda \ll \sqrt{D_p T \log(M)}$, the minimization of the VaR becomes equivalent to the minimization of the variance, and one recovers the Markowitz procedure explained in the previous paragraph in the case of Gaussian assets.

- If on the contrary $\Lambda \gg \sqrt{D_p T \log(M)}$, then the formulae established in the present section are valid *even when* $\mu > 2$.

Note that the growth of $\sqrt{\log(M)}$ with M is so slow that the Gaussian CLT is not of great help in the present case. The optimal portfolios in terms of the VaR are not those with the minimal variance, and vice versa.

3.2.3 'Exponential' assets

Suppose now that the distribution of price variations is a symmetric exponential around a zero mean value ($m_i = 0$):

$$P(\delta x_i) = \frac{\alpha_i}{2}\exp[-\alpha_i|\delta x_i|], \tag{3.54}$$

where α_i^{-1} gives the order of magnitude of the fluctuations of X_i (more precisely, $\sqrt{2}/\alpha_i$ is the RMS of the fluctuations.). The variations of the full portfolio **p**, defined as $\delta S = \sum_{i=1}^M p_i \delta x_i$, are distributed according to:

$$P(\delta S) = \frac{1}{2\pi}\int \frac{\exp(\mathrm{i}z\delta S)}{\prod_{i=1}^M [1 + (z p_i \alpha_i^{-1})^2]}\,\mathrm{d}z, \tag{3.55}$$

where we have used Eq. (1.50) and the fact that the Fourier transform of the exponential distribution Eq. (3.54) is given by:

$$\hat{P}(z) = \frac{1}{1 + (z\alpha^{-1})^2}. \tag{3.56}$$

Now, using the method of residues, it is simple to establish the following expression for $P(\delta S)$ (valid whenever the α_i/p_i are all different):

$$P(\delta S) = \frac{1}{2}\sum_{i=1}^{M} \frac{\alpha_i}{p_i} \frac{1}{\prod_{j\neq i}\left(1 - [(p_j\alpha_i)/(p_i\alpha_j)]^2\right)} \exp\left[-\frac{\alpha_i}{p_i}|\delta S|\right]. \tag{3.57}$$

The probability for extreme losses is thus equal to:

$$\mathcal{P}(\delta S < -\Lambda) \underset{\Lambda\to-\infty}{\simeq} \frac{1}{2\prod_{j\neq i^*}\left(1 - [(p_j\alpha^*)/\alpha_j]^2\right)} \exp[-\alpha^*\Lambda], \tag{3.58}$$

where α^* is equal to the smallest of all ratios α_i/p_i, and i^* the corresponding value of i. The order of magnitude of the extreme losses is therefore given by $1/\alpha^*$. This is then the quantity to be minimized in a value-at-risk context. This amounts to choosing the p_i's such that $\min_i\{\alpha_i/p_i\}$ is as large as possible.

This minimization problem can be solved using the following trick. One can

write formally that:

$$\frac{1}{\alpha^*} = \max\left\{\frac{p_i}{\alpha_i}\right\} = \lim_{\mu\to\infty}\left(\sum_{i=1}^{M}\frac{p_i^\mu}{\alpha_i^\mu}\right)^{\frac{1}{\mu}}. \tag{3.59}$$

This equality is true because in the limit $\mu \to \infty$, the largest term of the sum dominates over all the others. (The choice of the notation μ is on purpose: see below). For μ large but fixed, one can perform the minimization with respect to the p_i's, using a Lagrange multiplier to enforce normalization. One finds that:

$$p_i^{*\mu-1} \propto \alpha_i^\mu. \tag{3.60}$$

In the limit $\mu \to \infty$, and imposing $\sum_{i=1}^M p_i = 1$, one finally obtains:

$$p_i^* = \frac{\alpha_i}{\sum_{j=1}^M \alpha_j}, \tag{3.61}$$

which leads to $\alpha^* = \sum_{i=1}^M \alpha_i$. In this case, however, all α_i/p_i^* are equal to α^* and the result Eq. (3.57) must be slightly altered. However, the asymptotic exponential fall-off, governed by α^*, is still true (up to polynomial corrections: cf. Section 1.6.4). One again finds that if all the α_i's are comparable, the potential losses, measured through $1/\alpha^*$, are divided by a factor M.

Note that the optimal weights are such that $p_i^* \propto \alpha_i$ if one tries to minimize the probability of extreme losses, whereas one would have found $p_i^* \propto \alpha_i^2$ if the goal was to minimize the variance of the portfolio, corresponding to $\mu = 2$ (cf. Eq. (3.42)). In other words, this is an explicit example where one can see that minimizing the variance actually increases the value-at-risk.

Formally, as we have noticed in Section 1.3.4, the exponential distribution corresponds to the limit $\mu \to \infty$ of a power-law distribution: an exponential decay is indeed more rapid than any power-law. Technically, we have indeed established in this section that the minimization of extreme risks in the exponential case is identical to the one obtained in the previous section in the limit $\mu \to \infty$ (see Eq. (3.52)).

3.2.4 General case: optimal portfolio and VaR (*)

In all of the cases treated above, the optimal portfolio is found to be independent of the chosen loss level Λ. For example, in the case of assets with power-law tails, the minimization of the loss probability amounts to minimizing the tail amplitude A_p^μ, independently of Λ. This property is however not true in general, and the optimal portfolio does indeed depend on the risk level Λ, or, equivalently, on the temporal horizon over which risk must be 'tamed'. Let us for example consider the case where all assets are power-law distributed, but with a tail index μ_i that depends on the asset X_i. The probability that the portfolio $\mathbf{p}$ experiences of loss greater than Λ is given, for large values of Λ, by:[13]

$$\mathcal{P}(\delta S < -\Lambda) = \sum_{i=1}^M p_i^{\mu_i}\frac{A_i^{\mu_i}}{\Lambda^{\mu_i}}. \tag{3.62}$$

Looking for the set of p_i's which minimizes the above expression (without constraint on the average return) then leads to:

$$p_i^* = \frac{1}{Z}\left(\frac{\Lambda^{\mu_i}}{\mu_i A_i^\mu}\right)^{\frac{1}{\mu_i-1}} \qquad Z = \sum_{i=1}^M\left(\frac{\Lambda^{\mu_i}}{\mu_i A_i^\mu}\right)^{\frac{1}{\mu_i-1}}. \tag{3.63}$$

This example shows that in the general case, the weights p_i^* *explicitly depend on the risk level* Λ. If all the μ_i's are equal, Λ^{μ_i} factors out and disappears from the p_i's.

Another interesting case is that of weakly non-Gaussian assets, such that the first correction to the Gaussian distribution (proportional to the kurtosis κ_i of the asset X_i) is enough to describe faithfully the non-Gaussian effects. The variance of the full portfolio is given by $D_p = \sum_{i=1}^M p_i^2 D_i$ while the kurtosis is equal to: $\kappa_p = \sum_{i=1} p_i^4 D_i^2 \kappa_i / D_p^2$. The probability that the portfolio plummets by an amount larger than Λ is therefore given by:

$$\mathcal{P}(\delta S < -\Lambda) \simeq \mathcal{P}_{G>}\left(\frac{\Lambda}{\sqrt{D_p T}}\right) + \frac{\kappa_p}{4!}\, h\left(\frac{\Lambda}{\sqrt{D_p T}}\right), \tag{3.64}$$

where $\mathcal{P}_{G>}$ is related to the error function (cf. Section 1.6.3) and

$$h(u) = \frac{\exp(-u^2/2)}{\sqrt{2\pi}}(u^3 - 3u). \tag{3.65}$$

To first order in kurtosis, one thus finds that the optimal weights p_i^* (without fixing the average return) are given by:

$$p_i^* = \frac{\zeta'}{2D_i} - \frac{\kappa_i \zeta'^3}{D_i^3}\,\tilde{h}\left(\frac{\Lambda}{\sqrt{D_p T}}\right), \tag{3.66}$$

where $\tilde{h}$ is another function, positive for large arguments, and ζ' is fixed by the condition $\sum_{i=1}^M p_i = 1$. Hence, the optimal weights do depend on the risk level Λ, via the kurtosis of the distribution. Furthermore, as could be expected, the minimization of extreme risks leads to a reduction of the weights of the assets with a large kurtosis κ_i.

3.3 Portfolios of correlated assets

The aim of the previous section was to introduce, in a somewhat simplified context, the most important ideas underlying portfolio optimization, in a Gaussian world (where the variance is minimized) or in a non-Gaussian world (where

[13] The following expression is valid only when the *subleading corrections* to Eq. (3.47) can safely be neglected.

the quantity of interest is the value-at-risk). In reality, the fluctuations of the different assets are often strongly correlated (or anti-correlated). For example, an increase of short-term interest rates often leads to a drop in share prices. All the stocks of the New York Stock Exchange behave, to a certain extent, similarly. These correlations of course modify completely the composition of the optimal portfolios, and actually make diversification more difficult. In a sense, the number of effectively independent assets is decreased from the true number of assets M.

3.3.1 Correlated Gaussian fluctuations

Let us first consider the case where all the fluctuations δx_i of the assets X_i are Gaussian, but with arbitrary correlations. These correlations are described in terms of a (symmetric) correlation matrix C_{ij}, defined as:

$$C_{ij} = \langle \delta x_i \delta x_j \rangle - m_i m_j. \tag{3.67}$$

This means that the joint distribution of all the fluctuations $\delta x_1, \delta x_2, \ldots, \delta x_M$ is given by:

$$P(\delta x_1, \delta x_2, \ldots, \delta x_M) \propto \exp\left[-\frac{1}{2}\sum_{ij}(\delta x_j - m_j)(C^{-1})_{ij}(\delta x_j - m_j)\right], \tag{3.68}$$

where the proportionality factor is fixed by normalization and is equal to $1/\sqrt{(2\pi)^N \det \mathbf{C}}$, and $(C^{-1})_{ij}$ denotes the elements of the matrix inverse of $\mathbf{C}$.

An important property of correlated Gaussian variables is that they can be decomposed into a weighted sum of *independent* Gaussian variables e_a, of mean zero and variance equal to D_a:

$$\delta x_i = m_i + \sum_{a=1}^{M} O_{ia} e_a \qquad \langle e_a e_b \rangle = \delta_{a,b} D_a. \tag{3.69}$$

The $\{e_a\}$ are usually referred to as the 'explicative factors' (or principal components) for the asset fluctuations. They sometimes have a simple economic interpretation.

The coefficients O_{ia} give the weight of the factor e_a in the evolution of the asset X_i. These can be related to the correlation matrix C_{ij} by using the fact that the $\{e_a\}$'s are independent. This leads to:

$$C_{ij} = \sum_{a,b=1}^{M} O_{ia} O_{jb} \langle e_a e_b \rangle = \sum_{a=1}^{M} O_{ia} O_{ja} D_a, \tag{3.70}$$

or, seen as a matrix equality: $C = O\hat{D}O^\dagger$, where $O^\dagger$ denotes the matrix transposed of O and $\hat{D}$ the diagonal matrix obtained from the D_a's. This last expression shows

that the D_a's are the *eigenvalues* of the matrix C_{ij}, whereas O is the orthogonal matrix allowing one to go from the set of assets i's to the set of explicative factors e_a.

The fluctuations δS of the global portfolio $\mathbf{p}$ are then also Gaussian (since δS is a weighted sum of the Gaussian variables e_a), of mean $m_p = \sum_{i=1}^{M} p_i(m_i - m_0) + m_0$ and variance:

$$D_p = \sum_{i,j=1}^{M} p_i p_j C_{ij}. \tag{3.71}$$

The minimization of D_p for a fixed value of the average return m_p (and with the possibility of including the risk-free asset X_0) leads to an equation generalizing Eq. (3.38):

$$2\sum_{j=1}^{M} C_{ij} p_j^* = \zeta(m_i - m_0), \tag{3.72}$$

which can be inverted as:

$$p_i^* = \frac{\zeta}{2}\sum_{j=1}^{M} C_{ij}^{-1}(m_j - m_0). \tag{3.73}$$

This is Markowitz's classical result (cf. [Markowitz, Elton and Gruber]).

In the case where the risk-free asset is excluded, the minimum variance portfolio is given by:

$$p_i^* = \frac{1}{Z}\sum_{j=1}^{M} C_{ij}^{-1} \qquad Z = \sum_{i,j=1}^{M} C_{ij}^{-1}. \tag{3.74}$$

Actually, the decomposition, Eq. (3.69), shows that, provided one shifts to the basis where all assets are independent (through a linear combination of the original assets), all the results obtained above in the case where the correlations are absent (such as the existence of an efficient border, etc.) are still valid when correlations are present.

In the more general case of non-Gaussian assets of finite variance, the total variance of the portfolio is still given by: $\sum_{i,j=1}^{M} p_i p_j C_{ij}$, where C_{ij} is the correlation matrix. If the variance is an adequate measure of risk, the composition of the optimal portfolio is still given by Eqs (3.73) and (3.74). Let us however again emphasize that, as discussed in Section 2.7, the empirical determination of the correlation matrix C_{ij} is difficult, in particular when one is concerned with the small eigenvalues of this matrix and their corresponding eigenvectors.

The ideas developed in Section 2.7 can actually be used in practice to reduce the real risk of optimized portfolios. Since the eigenstates corresponding to the 'noise band' are not expected to contain real information, one should not distinguish the different eigenvalues

and eigenvectors in this sector. This amounts to replacing the restriction of the empirical correlation matrix to the noise band subspace by the identity matrix with a coefficient such that the trace of the matrix is conserved (i.e. suppressing the measurement broadening due to a finite observation time). This 'cleaned' correlation matrix, where the noise has been (at least partially) removed, is then used to construct an optimal portfolio. We have implemented this idea in practice as follows. Using the same data sets as above, the total available period of time has been divided into two equal sub-periods. We determine the correlation matrix using the first sub-period, 'clean' it, and construct the family of optimal portfolios and the corresponding efficient frontiers. Here we assume that the investor has perfect predictions on the future average returns m_i, i.e. we take for m_i the observed return on the next sub-period. The results are shown in Figure 3.8: one sees very clearly that using the empirical correlation matrix leads to a dramatic underestimation of the real risk, by over-investing in artificially low-risk eigenvectors. The risk of the optimized portfolio obtained using a cleaned correlation matrix is more reliable, although the real risk is always larger than the predicted one. This comes from the fact that any amount of uncertainty in the correlation matrix produces, through the very optimization procedure, a bias towards low-risk portfolios. This can be checked by permuting the two sub-periods: one then finds nearly identical efficient frontiers. (This is expected, since for large correlation matrices these frontiers should be self-averaging.) In other words, even if the cleaned correlation matrices are more stable in time than the empirical correlation matrices, they are not perfectly reflecting future correlations. This might be due to a combination of remaining noise and of a genuine time dependence in the structure of the meaningful correlations.

The CAPM and its limitations

Within the above framework, all optimal portfolios are proportional to one another, that is, they only differ through the choice of the factor ζ. Since the problem is linear, this means that the linear superposition of optimal portfolios is still optimal. If all the agents on the market choose their portfolio using this optimization scheme (with the same values for the average return and the correlation coefficients – clearly quite an absurd hypothesis), then the 'market portfolio' (i.e. the one obtained by taking all assets in proportion of their market capitalization) is an optimal portfolio. This remark is at the origin of the 'CAPM' (Capital Asset Pricing Model), which aims at relating the average return of an asset with its covariance with the 'market portfolio'. Actually, for any optimal portfolio $\mathbf{p}$, one can express $m_p - m_0$ in terms of the p^*, and use Eq. (3.73) to eliminate ζ, to obtain the following equality:

$$m_i - m_0 = \beta_i[m_p - m_0] \qquad \beta_i \equiv \frac{\langle(\delta x_i - m_i)(\delta S - m_p)\rangle}{\langle(\delta S - m_p)^2\rangle}. \tag{3.75}$$

The covariance coefficient β_i is often called the 'β' of asset i when $\mathbf{p}$ is the market portfolio.

This relation is however not true for other definitions of optimal portfolios. Let us define the generalized kurtosis K_{ijkl} that measures the first correction to Gaussian statistics, from the joint distribution of the asset fluctuations:

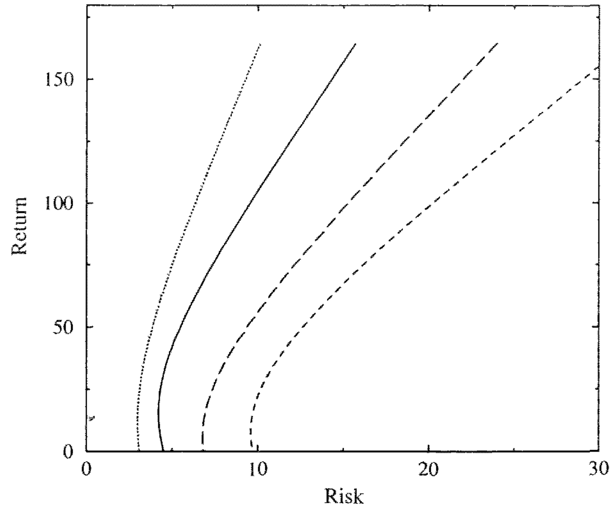

Fig. 3.8. Efficient frontiers from Markowitz optimization, in the return versus volatility plane. The leftmost dotted curve corresponds to the classical Markowitz case using the empirical correlation matrix. the rightmost short-dashed curve is the realization of the same portfolio in the second time period (the risk is underestimated by a factor of 3!). The central curves (plain and long-dashed) represent the case of a cleaned correlation matrix. The realized risk is now only a factor of 1.5 larger than the predicted risk.

$$P(\delta x_1, \delta x_2, \ldots, \delta x_M) =$$
$$\left(\frac{1}{2\pi}\right)^M \int\int\cdots\int \exp\left[-\mathrm{i}\sum_j z_j(\delta x_j - m_j)\right. \tag{3.76}$$
$$\left. -\frac{1}{2}\sum_{ij} z_i C_{ij} z_j + \frac{1}{4!}\sum_{ijkl} K_{ijkl} z_i z_j z_k z_l + \cdots\right]\prod_{j=1}^{M} \mathrm{d}z_j. \tag{3.77}$$

If one tries to minimize the probability that the loss is greater than a certain Λ, a generalization of the calculation presented above (cf. Eq. (3.66)) leads to:

$$p_i^* = \frac{\zeta'}{2}\sum_{j=1}^{M} C_{ij}^{-1}(m_j - m_0) - \zeta'^3 \tilde{h}\left(\frac{\Lambda}{\sqrt{D_p T}}\right)$$
$$\times \sum_{j,k,l=1}^{M}\ \sum_{j',k',l'=1}^{M} K_{ijkl} C_{jj'}^{-1} C_{kk'}^{-1} C_{ll'}^{-1}(m_{j'} - m_0)(m_{k'} - m_0)(m_{l'} - m_0). \tag{3.78}$$

where $\tilde{h}$ is a certain positive function. The important point here is that the level of risk Λ appears explicitly in the choice of the optimal portfolio. If the different operators choose

different levels of risk, their optimal portfolios are no longer related by a proportionality factor, and the above CAPM relation does not hold.

3.3.2 'Power-law' fluctuations (*)

The minimization of large risks also requires, as in the Gaussian case detailed above, the knowledge of the correlations between large events, and therefore an adapted measure of these correlations. Now, in the extreme case $\mu < 2$, the covariance (as the variance), is infinite. A suitable generalization of the covariance is therefore necessary. Even in the case $\mu > 2$, where this covariance is *a priori* finite, the value of this covariance is a mix of the correlations between large negative moves, large positive moves, and all the 'central' (i.e. not so large) events. Again, the definition of a 'tail covariance', directly sensitive to the large negative events, is needed. The aim of the present section is to define such a quantity, which is a natural generalization of the covariance for power-law distributions, much as the 'tail amplitude' is a generalization of the variance. In a second part, the minimization of the value-at-risk of a portfolio will be discussed.

'Tail covariance'

Let us again assume that the δx_i's are distributed according to:

$$P(\delta x_i) \underset{\delta x_i \to \pm\infty}{\simeq} \frac{\mu A_i^\mu}{|\delta x_i|^{1+\mu}}. \tag{3.79}$$

A natural way to describe the correlations between large events is to generalize the decomposition in independent factors used in the Gaussian case, Eq. (3.69) and to write:

$$\delta x_i = m_i + \sum_{a=1}^{M} O_{ia} e_a, \tag{3.80}$$

where the e_a are *independent* power-law random variables, the distribution of which is:

$$P(e_a) \underset{e_a \to \pm\infty}{\simeq} \frac{\mu A_a^\mu}{|e_a|^{1+\mu}}. \tag{3.81}$$

Since power-law variables are (asymptotically) stable under addition, the decomposition Eq. (3.80) indeed leads for all μ to correlated power-law variables δx_i.

The usual definition of the covariance is related to the average value of $\delta x_i \delta x_j$, which can in some cases be divergent (i.e. when $\mu < 2$). The idea is then to study directly the characteristic function of the product variable $\pi_{ij} = \delta x_i \delta x_j$.[14] The

[14] Other generalizations of the covariance have been proposed in the context of Lévy processes, such as the 'covariation' [Samorodnitsky and Taqqu].

justification is the following: if e_a and e_b are two independent power-law variables, their product π_{ab} is also a power-law variable (up to logarithmic corrections) with the same exponent μ (cf. Appendix C):

$$P(\pi_{ab}) \underset{\pi_{ab} \to \infty}{\simeq} \frac{\mu^2 (A_a^\mu A_b^\mu) \log(|\pi_{ab}|)}{|\pi_{ab}|^{1+\mu}} \qquad (a \neq b). \tag{3.82}$$

On the contrary, the quantity π_{aa} is distributed as a power-law *with an exponent* $\mu/2$ (cf. Appendix C):

$$P(\pi_{aa}) \underset{\pi_{aa} \to \infty}{\simeq} \frac{\mu A_a^\mu}{2|\pi_{aa}|^{1+\frac{\mu}{2}}}. \tag{3.83}$$

Hence, the variable π_{ij} gets both 'non-diagonal' contributions π_{ab} and 'diagonal' ones π_{aa}. For $\pi_{ij} \to \infty$, however, only the latter survive. Therefore π_{ij} is a power-law variable of exponent $\mu/2$, with a tail amplitude that we will note $A_{ij}^{\mu/2}$, and an asymmetry coefficient β_{ij} (see Section 1.3.3). Using the additivity of the tail amplitudes, and the results of Appendix C, one finds:

$$A_{ij}^{\mu/2} = \sum_{a=1}^{M} |O_{ia} O_{ja}|^{\frac{\mu}{2}} A_a^\mu, \tag{3.84}$$

and

$$C_{ij}^{\mu/2} \equiv \beta_{ij} A_{ij}^{\mu/2} = \sum_{a=1}^{M} \mathrm{sign}(O_{ia} O_{ja}) |O_{ia} O_{ja}|^{\frac{\mu}{2}} A_a^\mu. \tag{3.85}$$

In the limit $\mu = 2$, one sees that the quantity $C_{ij}^{\mu/2}$ reduces to the standard covariance, provided one identifies A_a^2 with the variance of the explicative factors D_a. This suggests that $C_{ij}^{\mu/2}$ is the suitable generalization of the covariance for power-law variables, which is constructed using extreme events only. The expression Eq. (3.85) furthermore shows that the matrix $\tilde{O}_{ia} = \mathrm{sign}(O_{ia})|O_{ia}|^{\mu/2}$ allows one to diagonalize the 'tail covariance matrix' $C_{ij}^{\mu/2}$;[15] its eigenvalues are given by the A_a^μ's.

In summary, the tail covariance matrix $C_{ij}^{\mu/2}$ is obtained by studying the asymptotic behaviour of the product variable $\delta x_i \delta x_j$, which is a power-law variable of exponent $\mu/2$. The product of its tail amplitude and of its asymmetry coefficient is our definition of the tail covariance $C_{ij}^{\mu/2}$.

[15] Note that $\tilde{O} = O$ for $\mu = 2$.

Optimal portfolio

It is now possible to find the optimal portfolio that minimizes the loss probability. The fluctuations of the portfolio $\mathbf{p}$ can indeed be written as:

$$\delta S \equiv \sum_{i=1}^{M} p_i \delta x_i = \sum_{a=1}^{M} \left(\sum_{i=1}^{M} p_i O_{ia} \right) e_a. \tag{3.86}$$

Due to our assumption that the e_a are symmetric power-law variables, δS is also a symmetric power-law variable with a tail amplitude given by:

$$A_p^\mu = \sum_{a=1}^{M} \left| \sum_{i=1}^{M} p_i O_{ia} \right|^\mu A_a^\mu. \tag{3.87}$$

In the simple case where one tries to minimize the tail amplitude A_p^μ without any constraint on the average return, one then finds:[16]

$$\sum_{a=1}^{M} O_{ia} A_a^\mu V_a^* = \frac{\zeta'}{\mu}, \tag{3.88}$$

where the vector $V_a^* = \mathrm{sign}(\sum_{j=1}^{M} O_{ja} p_j^*) |\sum_{j=1}^{M} O_{ja} p_j^*|^{\mu-1}$. Once the tail covariance matrix $C^{\mu/2}$ is known, the optimal weights p_i^* are thus determined as follows:

(i) The diagonalization of $C^{\mu/2}$ gives the rotation matrix $\tilde{O}$, and therefore one can construct the matrix $O = \mathrm{sign}(\tilde{O})|\tilde{O}|^{2/\mu}$ (understood for each element of the matrix), and the diagonal matrix A.
(ii) The matrix OA is then inverted, and applied to the vector $(\zeta'/\mu)\vec{1}$. This gives the vector $\vec{V}^* = \zeta'/\mu(OA)^{-1}\vec{1}$.
(iii) From the vector $\vec{V}^*$ one can then obtain the weights p_i^* by applying the matrix inverse of $O^\dagger$ to the vector $\mathrm{sign}(V^*)|V^*|^{1/(\mu-1)}$;
(iv) The Lagrange multiplier ζ' is then determined such as $\sum_{i=1}^{M} p_i^* = 1$.

Rather symbolically, one thus has:

$$\mathbf{p} = (O^\dagger)^{-1} \left(\frac{\zeta'}{\mu} (OA)^{-1} \vec{1} \right)^{\frac{1}{\mu-1}}. \tag{3.89}$$

In the case $\mu = 2$, one recovers the previous prescription, since in that case: $OA = OD = CO$, $(OA)^{-1} = O^{-1}C^{-1}$, and $O^\dagger = O^{-1}$, from which one gets:

$$\mathbf{p} = \frac{\zeta'}{2} O O^{-1} C^{-1} \vec{1} = \frac{\zeta'}{2} C^{-1} \vec{1}, \tag{3.90}$$

which coincides with Eq. (3.74).

[16] If one wants to impose a non-zero average return, one should replace ζ'/μ by $\zeta'/\mu + \zeta m_i/\mu$, where ζ is determined such as to give to m_p the required value.

The problem of the minimization of the value-at-risk in a strongly non-Gaussian world, where distributions behave asymptotically as power-laws, and in the presence of correlations between tail events, can thus be solved using a procedure rather similar to the one proposed by Markowitz for Gaussian assets.

3.4 Optimized trading (*)

In this section, we will discuss a slightly different problem of portfolio optimization: how to dynamically optimize, as a function of time, the number of shares of a given stock that one holds in order to minimize the risk for a fixed level of return. We shall actually encounter a similar problem in the next chapter on options when the question of the optimal hedging strategy will be addressed. In fact, much of the notations and techniques of the present section are borrowed from Chapter 4. The optimized strategy found below shows that in order to minimize the variance, the time-dependent part of the optimal strategy consists in selling when the price goes up and buying when it goes down. However, this strategy increases the probability of very large losses!

We will suppose that the trader holds a certain number of shares $\phi_n(x_n)$, where ϕ depends both on the (discrete) time $t_n = n\tau$, and on the price of the stock x_n. For the time being, we assume that the interest rates are negligible and that the change of wealth is given by:

$$\Delta W_X = \sum_{k=0}^{N-1} \phi_k(x_k) \delta x_k, \tag{3.91}$$

where $\delta x_k = x_{k+1} - x_k$. (See Sections 4.1 and 4.2 for a more detailed discussion of Eq. (3.91).) Let us define the gain $\mathcal{G} = \langle \Delta W_X \rangle$ as the average final wealth, and the risk $\mathcal{R}^2$ as the variance of the final wealth:

$$\mathcal{R}^2 = \langle \Delta W_X^2 \rangle - \mathcal{G}^2. \tag{3.92}$$

The question is then to find the optimal trading strategy $\phi_k^*(x_k)$, such that the risk $\mathcal{R}$ is minimized, for a given value of $\mathcal{G}$. Introducing a Lagrange multiplier ζ, one thus looks for the (functional) solution of the following equation:[17]

$$\frac{\delta}{\delta \phi_k(x)} \left[\mathcal{R}^2 - \zeta \mathcal{G}^2 \right] \Big|_{\phi_k = \phi_k^*} = 0. \tag{3.93}$$

Now, we further assume that the price increments δx_n are independent random variables of mean $m\tau$ and variance $D\tau$. Introducing the notation $P(x, k|x_0, 0)$ for the price to be equal to x at time $k\tau$, knowing that it is equal to x_0 at time $t = 0$,

[17] The following equation results from a functional minimization of the risk. See Section 4.4.3 for further details

one has:

$$\mathcal{G} = \sum_{k=0}^{N-1} \langle \phi_k(x_k) \rangle \langle \delta x_k \rangle \equiv m\tau \sum_{k=0}^{N-1} \int P(x, k|x_0, 0)\phi_k(x)\,\mathrm{d}x, \tag{3.94}$$

where the factorization of the average values holds because of the absence of correlations between the value of x_k (and thus that of ϕ_k), and the value of the increment δx_k. The calculation of $\langle \Delta W_X^2 \rangle$ is somewhat more involved, in particular because averages of the type $\langle \delta x_\ell \phi_k(x_k) \rangle$, with $k > \ell$ do appear: in this case one cannot neglect the correlations between δx_ℓ and x_k. Using the results of Appendix D, one finally finds:

$$\begin{aligned} \langle \Delta W_X^2 \rangle = D\tau \sum_{k=0}^{N-1} \int P(x, k|x_0, 0)\phi_k^2(x)\,\mathrm{d}x \\ + \quad m\tau \sum_{k=0}^{N-1}\sum_{\ell=0}^{k-1} \int\int P(x', \ell|x_0, 0)P(x, k|x', \ell)\phi_\ell(x')\phi_k(x)\, \frac{x - x'}{k - \ell}\,\mathrm{d}x\,\mathrm{d}x'. \end{aligned} \tag{3.95}$$

Taking the derivative of $P^2 - \zeta \mathcal{G}^2$ with respect to $\phi_k(x)$, one finds:

$$\begin{aligned} &2D\tau P(x, k|x_0, 0)\phi_k(x) - 2(1 + \zeta)m\tau P(x, k|x_0, 0)\mathcal{G} \\ &+ m\tau P(x, k|x_0, 0) \sum_{\ell=k+1}^{N-1} \int P(x', \ell|x, k)\phi_\ell(x')\frac{x' - x}{\ell - k}\,\mathrm{d}x' \\ &+ m\tau \sum_{\ell=0}^{k-1} \int P(x', \ell|x_0, 0)P(x, k|x', \ell)\phi_\ell(x')\frac{x - x'}{k - \ell}\,\mathrm{d}x'. \end{aligned} \tag{3.96}$$

Setting this expression to zero gives an implicit equation for the optimal strategy ϕ_k^*. A solution to this equation can be found, for m small, as a power series in m. Looking for a reasonable return means that $\mathcal{G}$ should be of order m. Therefore we set: $\mathcal{G} = \mathcal{G}_0 mT$, with $T = N\tau$, and expand ϕ^* and $1 + \zeta$ as:

$$\phi_k^* = \phi_k^0 + m\phi_k^1 + \cdots \quad \text{and} \quad \zeta = \frac{\zeta_0}{m^2} + \frac{\zeta_1}{m} + \cdots \tag{3.97}$$

Inserting these expressions in Eq. (3.96) leads, to zero-th order, to a time-independent strategy:

$$\phi_k^0(x) = \phi_0 = \frac{\zeta_0 \mathcal{G}_0 T}{D}. \tag{3.98}$$

The Lagrange multiplier ζ_0 is then fixed by the equation:

$$\mathcal{G} = \mathcal{G}_0 mT = mT\phi_0 \quad \text{leading to } \zeta_0 = \frac{D}{T} \text{ and } \phi_0 = \mathcal{G}_0. \tag{3.99}$$

To first order, the equation on ϕ_k^1 reads:

$$\begin{aligned} D\phi_k^1 &= \zeta_1 \mathcal{G}_0 - \frac{\phi_0}{2} \sum_{\ell=k+1}^{N-1} \int P(x', \ell|x, k)\frac{x' - x}{\ell - k}\,\mathrm{d}x' \\ &\quad - \frac{\phi_0}{2} \sum_{\ell=0}^{k-1} \int \frac{P(x', \ell|x_0, 0)P(x, k|x', \ell)}{P(x, k|x_0, 0)} \frac{x - x'}{k - \ell}\,\mathrm{d}x'. \end{aligned} \tag{3.100}$$

The second term on the right-hand side is of order m, and thus negligible to this order. Interestingly, the last term can be evaluated explicitly without any assumption on the detailed shape for the probability distribution of the price increments. Using the method of Appendix D, one can show that for independent increments:

$$\int (x' - x_0)P(x', \ell|x_0, 0)P(x, k|x', \ell)\,\mathrm{d}x' = \frac{\ell}{k}(x - x_0)P(x, k|x_0, 0). \tag{3.101}$$

Therefore, one finally finds:

$$\phi_k^1(x) = \frac{\zeta_1 \mathcal{G}_0 T}{D} - \frac{\phi_0}{2D}(x - x_0). \tag{3.102}$$

This equation shows that in order to minimize the risk *as measured by the variance*, a trader should sell stocks when their price increases, and buy more stocks when their price decreases, proportionally to $m(x - x_0)$. The value of ζ_1 is fixed such that:

$$\sum_{k=0}^{N-1} \int P(x, k|x_0, 0)\phi_k^1(x)\,\mathrm{d}x = 0, \tag{3.103}$$

which leads to $\zeta_1 = 0$ (plus order m corrections). However, it can be shown that this strategy increases the VaR. For example, if the increments are Gaussian, the left tail of the distribution of ΔW_X (corresponding to large losses), using the above strategy, is found to be exponential, and therefore much broader than the Gaussian expected for a time-independent strategy:

$$\mathcal{P}_<(\Delta W_X) \simeq_{\Delta W_X \to -\infty} \exp\left(-\frac{2|\Delta W_X|}{\mathcal{G}}\right). \tag{3.104}$$

3.5 Conclusion of the chapter

In a Gaussian world, all measures of risk are equivalent. Minimizing the variance or the probability of large losses (or the value-at-risk) lead to the same result, which is the family of optimal portfolios obtained by Markowitz. In the general case, however, the VaR minimization leads to portfolios which are different from those of Markowitz. These portfolios depend on the level of risk Λ (or on the time horizon

4

Futures and options: fundamental concepts

Les personnes non averties sont sujettes à se laisser induire en erreur.[1]

(Lord Raglan, 'Le tabou de l'inceste', quoted by Boris Vian in *L'automne à Pékin*.)

4.1 Introduction

4.1.1 Aim of the chapter

The aim of this chapter is to introduce the general theory of derivative pricing in a simple and intuitive, but rather unconventional, way. The usual presentation, which can be found in all the available books on the subject,[2] relies on particular models where it is possible to construct *riskless hedging strategies*, which replicate exactly the corresponding derivative product.[3] Since the risk is strictly zero, there is no ambiguity in the price of the derivative: it is equal to the cost of the hedging strategy. In the general case, however, these 'perfect' strategies do not exist. Not surprisingly for the layman, zero risk is the exception rather than the rule. Correspondingly, a suitable theory must include risk as an essential feature, which one would like to *minimize*. The present chapter thus aims at developing simple methods to obtain optimal strategies, residual risks, and prices of derivative products, which takes into account in an adequate way the peculiar statistical nature of financial markets, as described in Chapter 2.

4.1.2 Trading strategies and efficient markets

In the previous chapters, we have insisted on the fact that if the detailed prediction of future market moves is probably impossible, its statistical description is a reasonable and useful idea, at least as a first approximation. This approach only

1 Unwarned people may easily be fooled.
2 See e.g. [Hull, Wilmott, Baxter].
3 A hedging strategy is a trading strategy allowing one to reduce, and sometimes eliminate, the risk.

relies on a certain degree of stability (in time) in the way markets behave and the prices evolve.[4] Let us thus assume that one can determine (using a statistical analysis of past time series) the probability density $P(x, t|x_0, t_0)$, which gives the probability that the price of the asset X is equal to x (to within $\mathrm{d}x$) at time t, knowing that at a previous time t_0, the price was equal to x_0. As in previous chapters, we shall denote as $\langle \mathcal{O} \rangle$ the average (over the 'historical' probability distribution) of a certain observable $\mathcal{O}$:

$$\langle \mathcal{O}(x, t) \rangle \equiv \int P(x, t|x_0, 0)\mathcal{O}(x, t)\, \mathrm{d}x. \tag{4.1}$$

As we have shown in Chapter 2, the price fluctuations are somewhat correlated for small time intervals (a few minutes), but become rapidly uncorrelated (but not necessarily independent!) on longer time scales. In the following, we shall choose as our elementary time scale τ an interval a few times larger than the correlation time – say $\tau = 30$ min on liquid markets. We shall thus assume that the correlations of price increments on two different intervals of size τ are negligible.[5] When correlations are small, the information on future movements based on the study of past fluctuations is weak. In this case, no systematic trading strategy can be more profitable (on the long run) than holding the market index – of course, one can temporarily 'beat' the market through sheer luck. This property corresponds to the *efficient market hypothesis*.[6]

It is interesting to translate this property into more formal terms. Let us suppose that at time $t_n = n\tau$, an investor has a portfolio containing, in particular, a quantity $\phi_n(x_n)$ of the asset X, quantity which can depend on the price of the asset $x_n = x(t_n)$ at time t_n (this strategy could actually depend on the price of the asset for all previous times: $\phi_n(x_n, x_{n-1}, x_{n-2}, \ldots)$). Between t_n and t_{n+1}, the price of X varies by δx_n. This generates a profit (or a loss) for the investor equal to $\phi_n(x_n)\delta x_n$. Note that the change of wealth is *not* equal to $\delta(\phi x) = \phi\delta x + x\delta\phi$, since the second term only corresponds to converting some stocks in cash, or vice versa, but not to a real change of wealth. The wealth difference between time $t = 0$ and time

4 A weaker hypothesis is that the statistical 'texture' of the markets (i.e. the shape of the probability distributions) is stable over time, but that the *parameters* which fix the amplitude of the fluctuations can be time dependent – see Sections 1.7, 2.4 and 4.3.4.
5 The presence of small correlations on larger time scales is however difficult to exclude, in particular on the scale of several years, which might reflect economic cycles for example. Note also that the 'volatility' fluctuations exhibit long-range correlations – cf. Section 2.4.
6 The existence of successful systematic hedge funds, which have consistently produced returns higher than the average for several years, suggests that some sort of 'hidden' correlations do exist, at least on certain markets. But if they exist these correlations must be small and therefore are not relevant for our main concerns: risk control and option pricing.

$t_N = T = N\tau$, due to the trading of asset X is, for zero interest rates, equal to:

$$\Delta W_X = \sum_{n=0}^{N-1} \phi_n(x_n)\delta x_n. \tag{4.2}$$

Since the trading strategy $\phi_n(x_n)$ can only be chosen *before* the price actually changes, the absence of correlations means that the average value of ΔW_X (with respect to the historical probability distribution) reads:

$$\langle \Delta W_X \rangle = \sum_{n=0}^{N-1} \langle \phi_n(x_n) \rangle \langle \delta x_n \rangle \equiv \tilde{m}\tau \sum_{n=0}^{N-1} \langle x_n \phi_n(x_n) \rangle, \tag{4.3}$$

where we have introduced the average return $\tilde{m}$ of the asset X, defined as:

$$\delta x_n \equiv \eta_n x_n \qquad \tilde{m}\tau = \langle \eta_n \rangle. \tag{4.4}$$

The above equation (4.3) thus means that the average value of the profit is fixed by the average return of the asset, weighted by the level of investment across the considered period. We shall often use, in the following, an additive model (more adapted on short time scales, cf. Section 2.2.1) where δx_n is rather written as $\delta x_n = \eta_n x_0$. Correspondingly, the average return over the time interval τ reads: $m_1 = \langle \delta x \rangle = \tilde{m}\tau x_0$. This approximation is usually justified as long as the total time interval T corresponds to a small (average) relative increase of price: $\tilde{m}T \ll 1$. We will often denote as m the average return *per unit time*: $m = m_1/\tau$.

Trading in the presence of temporal correlations

*It is interesting to investigate the case where correlations are not zero. For simplicity, we shall assume that the fluctuations δx_n are stationary Gaussian variables of zero mean ($m_1 = 0$). The correlation function is then given by $\langle \delta x_n \delta x_k \rangle \equiv C_{nk}$. The C_{nk}^{-1}'s are the elements of the matrix inverse of **C**. If one knows the sequence of past increments $\delta x_0, \ldots, \delta x_{n-1}$, the distribution of the next δx_n conditioned to such an observation is simply given by:*

$$P(\delta x_n) = \mathcal{N} \exp -\frac{C_{nn}^{-1}}{2}[\delta x_n - m_n]^2, \tag{4.5}$$

$$m_n \equiv -\frac{\sum_{i=0}^{n-1} C_{in}^{-1}\delta x_i}{C_{nn}^{-1}}, \tag{4.6}$$

where $\mathcal{N}$ is a normalization factor, and m_n the mean of δx_n conditioned to the past, which is non-zero precisely because some correlations are present.[7] *A simple strategy which exploits these correlations is to choose:*

$$\phi_n(x_n, x_{n-1}, \ldots) = \mathrm{sign}(m_n), \tag{4.7}$$

[7] The notation m_n has already been used in Chapter 1 with a different meaning. Note also that a general formula exists for the distribution of δx_{n+k} for all $k \geq 0$, and can be found in books on optimal filter theory, see references.

which means that one buys (resp. sells) one stock if the expected next increment is positive (resp. negative). The average profit is then obviously given by $\langle |m_n| \rangle > 0$. We further assume that the correlations are short ranged, such that only C_{nn}^{-1} and C_{nn-1}^{-1} are non-zero. The unit time interval is then the above correlation time τ. If this strategy is used during the time $T = N\tau$, the average profit is given by:

$$\langle \Delta W_X \rangle = Na\langle |\delta x| \rangle \text{ where } a \equiv \left| \frac{C_{nn-1}^{-1}}{C_{nn}^{-1}} \right|, \tag{4.8}$$

(a does not depend on n for a stationary process). With typical values, $\tau = 30$ min, $\langle |\delta x| \rangle \simeq 10^{-3}\, x_0$, and a of about 0.1 (cf. Section 2.2.2), the average profit would reach 50% annual![8] *Hence, the presence of correlations (even rather weak) allows in principle one to make rather substantial gains. However, one should remember that some transaction costs are always present in some form or other (for example the bid-ask spread is a form of transaction cost). Since our assumption is that the correlation time is equal to τ, it means that the frequency with which our trader has to 'flip' his strategy (i.e. $\phi \to -\phi$) is τ^{-1}. One should thus subtract from Eq. (4.8) a term on the order of $-N\nu x_0$, where ν represents the fractional cost of a transaction. A very small value of the order of 10^{-4} is thus enough to cancel completely the profit resulting from the observed correlations on the markets. Interestingly enough, the 'basis point' (10^{-4}) is precisely the order of magnitude of the friction faced by traders with the most direct access to the markets. More generally, transaction costs allow the existence of correlations, and thus of 'theoretical' inefficiencies of the markets. The strength of the 'allowed' correlations on a certain time T is on the order of the transaction costs ν divided by the volatility of the asset on the scale of T.*

At this stage, it would appear that devising a 'theory' for trading is useless: in the absence of correlations, speculating on stock markets is tantamount to playing the roulette (the zero playing the role of transaction costs!). In fact, as will be discussed in the present chapter, the existence of riskless assets such as bonds, which yield a known return, and the emergence of more complex financial products such as futures contracts or options, demand an adapted theory for pricing and hedging. This theory turns out to be predictive, even in the complete absence of temporal correlations.

4.2 Futures and forwards

4.2.1 Setting the stage

Before turning to the rather complex case of options, we shall first focus on the very simple case of forward contracts, which allows us to define a certain number of notions (such as arbitrage and hedging) and notations. A forward contract F amounts to buying or selling today an asset X (the 'underlying') with a delivery

[8] A strategy that allows one to generate consistent abnormal returns with zero or minimal risk is called an 'arbitrage opportunity'.

date $T = N\tau$ in the future.[9] What is the price $\mathcal{F}$ of this contract, knowing that it must be paid at the date of expiry?[10] The naive answer that first comes to mind is a 'fair game' condition: the price must be adjusted such that, *on average*, the two parties involved in the contract fall even on the day of expiry. For example, taking the point of view of the writer of the contract (who sells the forward), the wealth balance associated with the forward reads:

$$\Delta W_F = \mathcal{F} - x(T). \tag{4.9}$$

This actually assumes that the writer has not considered the possibility of simultaneously trading the underlying stock X to reduce his risk, which of course he should do, see below.

Under this assumption, the fair game condition amounts to set $\langle \Delta W_F \rangle = 0$, which gives for the forward price:

$$\mathcal{F}_B = \langle x(T) \rangle \equiv \int x P(x, T|x_0, 0)\, \mathrm{d}x, \tag{4.10}$$

if the price of X is x_0 at time $t = 0$. This price, that can we shall call the 'Bachelier' price, is not satisfactory since the seller takes the risk that the price at expiry $x(T)$ ends up far above $\langle x(T) \rangle$, which could prompt him to increase his price above $\mathcal{F}_B$.

Actually, the Bachelier price $\mathcal{F}_B$ is not related to the market price for a simple reason: the seller of the contract can suppress his risk completely if he buys now the underlying asset X at price x_0 and waits for the expiry date. However, this strategy is costly: the amount of cash frozen during that period does not yield the riskless interest rate. The cost of the strategy is thus $x_0 e^{rT}$, where r is the interest rate per unit time. From the view point of the buyer, it would certainly be absurd to pay more than $x_0 e^{rT}$, which is the cost of borrowing the cash needed to pay the asset right away. The only viable price for the forward contract is thus $\mathcal{F} = x_0 e^{rT} \neq \mathcal{F}_B$, and is, in this particular case, completely unrelated to the potential moves of the asset X!

An elementary argument thus allows one to know the price of the forward contract and to follow a perfect hedging strategy: buy a quantity $\phi = 1$ of the underlying asset during the whole life of the contract. The aim of next paragraph is to establish this rather trivial result, which has not required any maths, in a much more sophisticated way. The importance of this procedure is that one needs to learn how to write down a proper wealth balance in order to price more complex derivative products such as options, on which we shall focus in the next sections.

[9] In practice 'futures' contracts are more common than forwards. While forwards are over-the-counter contracts, futures are traded on organized markets. For forwards there are typically no payments from either side before the expiration date whereas futures are marked-to-market and compensated every day, meaning that payments are made daily by one side or the other to bring the value of the contract back to zero.

[10] Note that if the contract was to be paid now, its price would be exactly equal to that of the underlying asset (barring the risk of delivery default).

4.2.2 Global financial balance

Let us write a general financial balance which takes into account the trading strategy of the the underlying asset X. The difficulty lies in the fact that the amount $\phi_n x_n$ which is invested in the asset X rather than in bonds is effectively costly: one 'misses' the risk-free interest rate. Furthermore, this loss cumulates in time. It is not *a priori* obvious how to write down the correct balance. Suppose that only two assets have to be considered: the risky asset X, and a bond B. The whole capital W_n at time $t_n = n\tau$ is shared between these two assets:

$$W_n \equiv \phi_n x_n + B_n. \tag{4.11}$$

The time evolution of W_n is due both to the change of price of the asset X, and to the fact that the bond yields a known return through the interest rate r:

$$W_{n+1} - W_n = \phi_n (x_{n+1} - x_n) + B_n \rho, \qquad \rho = r\tau. \tag{4.12}$$

On the other hand, the amount of capital in bonds evolves both mechanically, through the effect of the interest rate $(+B_n\rho)$, but also because the quantity of stock does change in time $(\phi_n \to \phi_{n+1})$, and causes a flow of money from bonds to stock or vice versa. Hence:

$$B_{n+1} - B_n = B_n \rho - x_{n+1}(\phi_{n+1} - \phi_n). \tag{4.13}$$

Note that Eq. (4.11) is obviously compatible with the following two equations. The solution of the second equation reads:

$$B_n = (1+\rho)^n B_0 - \sum_{k=1}^{n} x_k (\phi_k - \phi_{k-1})(1+\rho)^{n-k}. \tag{4.14}$$

Plugging this result in Eq. (4.11), and renaming $k - 1 \to k$ in the second part of the sum, one finally ends up with the following central result for W_n:

$$W_n = W_0 (1+\rho)^n + \sum_{k=0}^{n-1} \psi_k^n (x_{k+1} - x_k - \rho x_k), \tag{4.15}$$

with $\psi_k^n \equiv \phi_k (1+\rho)^{n-k-1}$. This last expression has an intuitive meaning: the gain or loss incurred between time k and $k+1$ must include the cost of the hedging strategy $-\rho x_k$; furthermore, this gain or loss must be forwarded up to time n through interest rate effects, hence the extra factor $(1+\rho)^{n-k-1}$. Another

useful way to write and understand Eq. (4.15) is to introduce the discounted prices $\tilde{x}_k \equiv x_k(1+\rho)^{-k}$. One then has:

$$W_n = (1+\rho)^n \left(W_0 + \sum_{k=0}^{n-1} \phi_k(\tilde{x}_{k+1} - \tilde{x}_k) \right). \tag{4.16}$$

The effect of interest rates can be thought of as an erosion of the value of the money itself. The discounted prices $\tilde{x}_k$ are therefore the 'true' prices, and the difference $\tilde{x}_{k+1} - \tilde{x}_k$ the 'true' change of wealth. The overall factor $(1+\rho)^n$ then converts this true wealth into the current value of the money.

The global balance associated with the forward contract contains two further terms: the price of the forward $\mathcal{F}$ that one wishes to determine, and the price of the underlying asset at the delivery date. Hence, finally:

$$W_N = \mathcal{F} - x_N + (1+\rho)^N \left(W_0 + \sum_{k=0}^{N-1} \phi_k(\tilde{x}_{k+1} - \tilde{x}_k) \right). \tag{4.17}$$

Since by identity $\tilde{x}_N = \sum_{k=0}^{N-1}(\tilde{x}_{k+1} - \tilde{x}_k) + x_0$, this last expression can also be written as:

$$W_N = \mathcal{F} + (1+\rho)^N \left(W_0 - x_0 + \sum_{k=0}^{N-1}(\phi_k - 1)(\tilde{x}_{k+1} - \tilde{x}_k) \right). \tag{4.18}$$

4.2.3 Riskless hedge

In this last formula, all the randomness, the uncertainty on the future evolution of the prices, only appears in the last term. But if one chooses ϕ_k to be identically equal to one, then the global balance is completely independent of the evolution of the stock price. The final result is *not* random, and reads:

$$W_N = \mathcal{F} + (1+\rho)^N(W_0 - x_0). \tag{4.19}$$

Now, the wealth of the writer of the forward contract at time $T = N\tau$ cannot be, with certitude, greater (or less) than the one he would have had if the contract had not been signed, i.e. $W_0(1+\rho)^N$. If this was the case, one of the two parties would be losing money in a totally predictable fashion (since Eq. (4.19) does not contain any unknown term). Since any reasonable participant is reluctant to give away his money with no hope of return, one concludes that the forward price is indeed given by:

$$\mathcal{F} = x_0(1+\rho)^N \simeq x_0 e^{rT} \neq \mathcal{F}_B, \tag{4.20}$$

which does not rely on any statistical information on the price of X!

Dividends

In the case of a stock that pays a constant dividend rate $\delta = d\tau$ per interval of time τ, the global wealth balance is obviously changed into:

$$W_N = \mathcal{F} - x_N + W_0(1+\rho)^N + \sum_{k=0}^{N-1} \psi_k^N(x_{k+1} - x_k + (\delta - \rho)x_k). \tag{4.21}$$

It is easy to redo the above calculations in this case. One finds that the riskless strategy is now to hold:

$$\phi_k \equiv \frac{(1+\rho-\delta)^{N-k-1}}{(1+\rho)^{N-k}} \tag{4.22}$$

stocks. The wealth balance breaks even if the price of the forward is set to:

$$\mathcal{F} = x_0(1+\rho-\delta)^N \simeq x_0 e^{(r-d)T}, \tag{4.23}$$

which again could have been obtained using a simple no arbitrage argument of the type presented below.

Variable interest rates

In reality, the interest rate is not constant in time but rather also varies randomly. More precisely, as explained in Section 2.6, at any instant of time the whole interest rate curve for different maturities is known, but evolves with time. The generalization of the global balance, as given by formula, Eq. (4.17), depends on the maturity of the bonds included in the portfolio, and thus on the whole interest rate curve. Assuming that only short-term bonds are included, yielding the (time-dependent) 'spot' rate ρ_k, one has:

$$\begin{aligned} W_N &= \mathcal{F} - x_N + W_0 \prod_{k=0}^{N-1}(1+\rho_k) \\ &\quad + \sum_{k=0}^{N-1} \psi_k^N(x_{k+1} - x_k - \rho_k x_k), \end{aligned} \tag{4.24}$$

with: $\psi_k^N \equiv \phi_k \prod_{\ell=k+1}^{N-1}(1+\rho_\ell)$. It is again quite obvious that holding a quantity $\phi_k \equiv 1$ of the underlying asset leads to zero risk in the sense that the fluctuations of X disappear. However, the above strategy is not immune to interest rate risk. The interesting complexity of interest rate problems (such as the pricing and hedging of interest rate derivatives) comes from the fact that one may choose bonds of arbitrary maturity to construct the hedging strategy. In the present case, one may take a bond of maturity equal to that of the forward. In this case, risk disappears entirely, and the price of the forward reads:

$$\mathcal{F} = \frac{x_0}{B(0,N)}, \tag{4.25}$$

where $B(0,N)$ stands for the value, at time 0, of the bond maturing at time N.

4.2.4 *Conclusion: global balance and arbitrage*

From the simple example of forward contracts, one should bear in mind the following points, which are the key concepts underlying the derivative pricing theory as presented in this book. After writing down the *complete* financial balance, taking into account the trading of all assets used to cover the risk, it is quite natural (at least from the view point of the writer of the contract) to determine the trading strategy for all the hedging assets so as to minimize the risk associated to the contract. After doing so, a reference price is obtained by demanding that the global balance is zero on average, corresponding to a fair price from the point of view of both parties. In certain cases (such as the forward contracts described above), the minimum risk is zero and the true market price cannot differ from the fair price, or else arbitrage would be possible. On the example of forward contracts, the price, Eq. (4.20), indeed corresponds to the *absence of arbitrage opportunities* (AAO), that is, of riskless profit. Suppose for example that the price of the forward is *below* $\mathcal{F} = x_0(1+\rho)^N$. One can then sell the underlying asset now at price x_0 and simultaneously buy the forward at a price $\mathcal{F}' < \mathcal{F}$, that must be paid for on the delivery date. The cash x_0 is used to buy bonds with a yield rate ρ. On the expiry date, the forward contract is used to buy back the stock and close the position. The wealth of the trader is then $x_0(1+\rho)^N - \mathcal{F}'$, which is positive under our assumption – and furthermore fully determined at time zero: there is profit, but no risk. Similarly, a price $\mathcal{F}' > \mathcal{F}$ would also lead to an opportunity of arbitrage. More generally, if the hedging strategy is perfect, this means that the final wealth is known in advance. Thus, increasing the price as compared to the fair price leads to a riskless profit for the seller of the contract, and vice versa. This AAO principle is at the heart of most derivative pricing theories currently used. Unfortunately, this principle cannot be used in the general case, where the minimal risk is non-zero, or when transaction costs are present (and absorb the potential profit, see the discussion in Section 4.1.2 above). When the risk is non-zero, there exists a fundamental ambiguity in the price, since one should expect that a risk premium is added to the fair price (for example, as a bid-ask spread). This risk premium depends both on the risk-averseness of the market maker, but also on the liquidity of the derivative market: if the price asked by one market maker is too high, less greedy market makers will make the deal. This mechanism does however not operate for 'over-the-counter' operations (OTC, that is between two individual parties, as opposed to through an organized market). We shall come back to this important discussion in Section 4.6 below.

Let us emphasize that the proper accounting of all financial elements in the wealth balance is crucial to obtain the correct fair price. For example, we have seen above that if one forgets the term corresponding to the trading of the underlying stock, one ends up with the intuitive, but wrong, Bachelier price, Eq. (4.10).

4.3 Options: definition and valuation

4.3.1 *Setting the stage*

A buy option (or 'call' option) is nothing but an insurance policy, protecting the owner against the potential increase of the price of a given asset, which he will need to buy in the future. The call option provides to its owner the certainty of not paying the asset more than a certain price. Symmetrically, a 'put' option protects against drawdowns, by insuring to the owner a minimum value for his stock.

More precisely, in the case of a so-called 'European' option,[11] the contract is such that at a given date in the future (the 'expiry date' or 'maturity') $t = T$, the owner of the option will not pay the asset more than x_s (the 'exercise price', or 'strike' price): the possible difference between the market price at time T, $x(T)$ and x_s is taken care of by the writer of the option. Knowing that the price of the underlying asset is x_0 now (i.e. at $t = 0$), what is the price (or 'premium') $\mathcal{C}$ of the call? What is the optimal hedging strategy that the writer of the option should follow in order to minimize his risk?

The very first scientific theory of option pricing dates back to Bachelier in 1900. His proposal was, following a fair game argument, that the option price should equal the average value of the pay-off of the contract at expiry. Bachelier calculated this average by assuming the price increments δx_n to be independent Gaussian random variables, which leads to the formula, Eq. (4.43) below. However, Bachelier did not discuss the possibility of hedging, and therefore did not include in his wealth balance the term corresponding to the trading strategy that we have discussed in the previous section. As we have seen, this is precisely the term responsible for the difference between the forward 'Bachelier price' $\mathcal{F}_B$ (cf. Eq. (4.10)) and the true price, Eq. (4.20). The problem of the optimal trading strategy must thus, in principle, be solved before one can fix the price of the option.[12] This is the problem that was solved by Black and Scholes in 1973, when they showed that for a continuous-time Gaussian process, there exists a perfect strategy, in the sense that the risk associated to writing an option is *strictly zero*, as is the case for forward contracts. The determination of this perfect hedging strategy allows one to fix completely the price of the option using an AAO argument. Unfortunately, as repeatedly discussed below, this ideal strategy only exists in

[11] Many other types of options exist: 'American', 'Asian', 'Lookback', 'Digitals', etc., see [Wilmott]. We will discuss some of them in Chapter 5.

[12] In practice, however, the influence of the trading strategy on the price (but not on the risk!) is quite small, see below.

a continuous-time, Gaussian world,[13] that is, if the market correlation time was infinitely short, and if no discontinuities in the market price were allowed – both assumptions rather remote from reality. The hedging strategy cannot, in general, be perfect. However, an *optimal* hedging strategy can always be found, for which the risk is minimal (cf. Section 4.4). This optimal strategy thus allows one to calculate the fair price of the option and the associated residual risk. One should nevertheless bear in mind that, as emphasized in Sections 4.2.4 and 4.6, there is no such thing as a unique option price whenever the risk is non-zero.

Let us now discuss these ideas more precisely. Following the method and notations introduced in the previous section, we write the global wealth balance for the writer of an option between time $t = 0$ and $t = T$ as:

$$W_N = [W_0 + \mathcal{C}](1+\rho)^N - \max(x_N - x_s, 0) + \sum_{k=0}^{N-1} \psi_k^N (x_{k+1} - x_k - \rho x_k), \tag{4.26}$$

which reflects the fact that:

- The premium $\mathcal{C}$ is paid immediately (i.e. at time $t = 0$).
- The writer of the option incurs a *loss* $x_N - x_s$ only if the option is exercised ($x_N > x_s$).
- The hedging strategy requires that the writer convert a certain amount of bonds into the underlying asset, as was discussed before Eq. (4.17).

A crucial difference with forward contracts comes from the *non-linear* nature of the pay-off, which is equal, in the case of a European option, to $\mathcal{Y}(x_N) \equiv \max(x_N - x_s, 0)$. This must be contrasted with the forward pay-off, which is linear (and equal to x_N). It is ultimately the non-linearity of the pay-off which, combined with the non-Gaussian nature of the fluctuations, leads to a non-zero residual risk.

An equation for the call price $\mathcal{C}$ is obtained by requiring that the excess return due to writing the option, $\Delta W = W_N - W_0(1+\rho)^N$, is zero on average:

$$(1+\rho)^N \mathcal{C} = \left[\langle \max(x_N - x_s, 0) \rangle - \sum_{k=0}^{N-1} \langle \psi_k^N (x_{k+1} - x_k - \rho x_k) \rangle \right]. \tag{4.27}$$

This price therefore depends, in principle, on the strategy $\psi_k^N = \phi_k^N (1+\rho)^{N-k-1}$. This price corresponds to the fair price, to which a risk premium will in general be added (for example in the form of a bid-ask spread).

In the rather common case where the underlying asset is not a stock but a forward

[13] A property also shared by a discrete binomial evolution of prices, where at each time step, the price increment δx can only take two values, see Appendix E. However, the risk is non-zero as soon as the number of possible price changes exceeds two.

on the stock, the hedging strategy is less costly since only a small fraction f of the value of the stock is required as a deposit. In the case where $f = 0$, the wealth balance appears to take a simpler form, since the interest rate is not lost while trading the underlying forward contract:

$$\mathcal{C} = (1+\rho)^{-N} \left[\langle \max(\mathcal{F}_N - x_s, 0) \rangle - \sum_{k=0}^{N-1} \langle \phi_k (\mathcal{F}_{k+1} - \mathcal{F}_k) \rangle \right]. \tag{4.28}$$

However, one can check that if one expresses the price of the forward in terms of the underlying stock, Eqs (4.27) and (4.28) are actually identical. (Note in particular that $\mathcal{F}_N = x_N$.)

4.3.2 Orders of magnitude

Let us suppose that the maturity of the option is such that non-Gaussian 'tail' effects can be neglected ($T > T^*$, cf. Sections 1.6.3, 1.6.5, 2.3), so that the distribution of the terminal price $x(T)$ can be approximated by a Gaussian of mean mT and variance $DT = \sigma^2 x_0^2 T$.[14] If the average trend is small compared to the RMS $\sqrt{DT}$, a direct calculation for 'at-the-money' options gives:[15]

$$\begin{aligned} \langle \max(x(T) - x_s, 0) \rangle &= \int_{x_s}^{\infty} \frac{x - x_s}{\sqrt{2\pi DT}} \exp\left(-\frac{(x - x_0 - mT)^2}{2DT} \right) \mathrm{d}x \\ &\simeq \sqrt{\frac{DT}{2\pi}} + \frac{mT}{2} + O\left(\sqrt{\frac{m^4 T^3}{D}} \right). \end{aligned} \tag{4.29}$$

Taking $T = 100$ days, a daily volatility of $\sigma = 1\%$, an annual return of $m = 5\%$, and a stock such that $x_0 = 100$ points, one gets:

$$\sqrt{\frac{DT}{2\pi}} \simeq 4 \text{ points} \qquad \frac{mT}{2} \simeq 0.67 \text{ points}. \tag{4.30}$$

In fact, the effect of a non-zero average return of the stock is much less than the above estimation would suggest. The reason is that one must also take into account the last term of the balance equation, Eq. (4.27), which comes from the hedging strategy. This term corrects the price by an amount $-\langle \phi \rangle mT$. Now, as we shall see later, the optimal strategy for an at-the-money option is to hold, on average $\langle \phi \rangle = \frac{1}{2}$ stocks per option. Hence, this term precisely compensates the increase (equal to $mT/2$) of the average pay-off, $\langle \max(x(T) - x_s, 0) \rangle$. This strange compensation is actually *exact* in the Black–Scholes model (where the price of

[14] We again neglect the difference between Gaussian and log-normal statistics in the following order of magnitude estimate. See Sections 1.3.2, 2.2.1, Eq. (4.43) and Figure 4.1 below.

[15] An option is called 'at-the-money' if the strike price is equal to the current price of the underlying stock ($x_s = x_0$), 'out-of-the-money' if $x_s > x_0$ and 'in-the-money' if $x_s < x_0$.

the option turns out to be completely independent of the value of m) but only approximate in more general cases. However, it is a very good first approximation to neglect the dependence of the option price in the average return of the stock: we shall come back to this point in detail in Section 4.5.

The interest rate appears in two different places in the balance equation, Eq. (4.27): in front of the call premium $\mathcal{C}$, and in the cost of the trading strategy. For $\rho = r\tau$ corresponding to 5% per year, the factor $(1+\rho)^N$ corrects the option price by a very small amount, which, for $T = 100$ days, is equal to 0.06 points. However, the cost of the trading strategy is not negligible. Its order of magnitude is $\sim \langle\phi\rangle x_0 r T$: in the above numerical example, it corresponds to a price increase of $\frac{2}{3}$ of a point (16% of the option price), see Section 5.1.

4.3.3 Quantitative analysis – option price

We shall assume in the following that the price increments can be written as:

$$x_{k+1} - x_k = \rho x_k + \delta x_k, \tag{4.31}$$

where δx_k is a random variable having the characteristics discussed in Chapter 2, and a mean value equal to $\langle \delta x_k \rangle = m_1 \equiv m\tau$. The above order of magnitude estimate suggests that the influence of a non-zero value of m leads to small corrections in comparison with the potential fluctuations of the stock price. We shall thus temporarily set $m = 0$ to simplify the following discussion, and come back to the corrections brought about by a non-zero value of m in Section 4.5.

Since the hedging strategy ψ_k is obviously determined *before* the next random change of price δx_k, these two quantities are uncorrelated, and one has:

$$\langle \psi_k \delta x_k \rangle = \langle \psi_k \rangle \langle \delta x_k \rangle = 0 \qquad (m = 0). \tag{4.32}$$

In this case, the hedging strategy disappears from the price of the option, which is given by the following 'Bachelier'-like formula, generalized to the case where the increments are not necessarily Gaussian:

$$\begin{aligned} \mathcal{C} &= (1+\rho)^{-N} \langle \max(x_N - x_s, 0) \rangle \\ &\equiv (1+\rho)^{-N} \int_{x_s}^{\infty} (x - x_s) P(x, N|x_0, 0)\, dx. \end{aligned} \tag{4.33}$$

In order to use Eq. (4.33) concretely, one needs to specify a model for price increments. In order to recover the classical model of Black and Scholes, let us first assume that the *relative* returns are iid random variables and write $\delta x_k \equiv \eta_k x_k$, with $\eta_k \ll 1$. If one knows (from empirical observation) the distribution $P_1(\eta_k)$ of returns over the elementary time scale τ, one can easily reconstruct (using the independence of the returns) the distribution $P(x, N|x_0, 0)$ needed to compute $\mathcal{C}$. After changing variables to $x \to x_0(1+\rho)^N e^y$, the formula, Eq. (4.33), is transformed into:

$$\mathcal{C} = x_0 \int_{y_s}^{\infty} (e^y - e^{y_s}) P_N(y)\, dy, \tag{4.34}$$

where $y_s \equiv \log(x_s/x_0[1+\rho]^N)$ and $P_N(y) \equiv P(y, N|0, 0)$. Note that y_s involves the ratio of the strike price to the *forward* price of the underlying asset, $x_0[1+\rho]^N$.

Setting $x_k = x_0(1+\rho)^k e^{y_k}$, the evolution of the y_k's is given by:

$$y_{k+1} - y_k \simeq \frac{\eta_k}{1+\rho} - \frac{\eta_k^2}{2} \qquad y_0 = 0, \tag{4.35}$$

where third-order terms ($\eta^3, \eta^2\rho, \ldots$) have been neglected. The distribution of the quantity $y_N = \sum_{k=0}^{N-1}[(\eta_k/(1+\rho)) - (\eta_k^2/2)]$ is then obtained, in Fourier space, as:

$$\hat{P}_N(z) = [\tilde{P}_1(z)]^N, \tag{4.36}$$

where we have defined, in the right-hand side, a modified Fourier transform:

$$\tilde{P}_1(z) = \int P_1(\eta) \exp\left[iz\left(\frac{\eta}{1+\rho} - \frac{\eta^2}{2}\right)\right] d\eta. \tag{4.37}$$

The Black and Scholes limit

We can now examine the Black–Scholes limit, where $P_1(\eta)$ is a Gaussian of zero mean and RMS equal to $\sigma_1 = \sigma\sqrt{\tau}$. Using the above Eqs (4.36) and (4.37) one finds, for N large:[16]

$$P_N(y) = \frac{1}{\sqrt{2\pi N\sigma_1^2}} \exp\left(-\frac{(y + N\sigma_1^2/2)^2}{2N\sigma_1^2}\right). \tag{4.38}$$

The Black–Scholes model also corresponds to the limit where the elementary time interval τ goes to zero, in which case one of course has $N = T/\tau \to \infty$. As discussed in Chapter 2, this limit is not very realistic since any transaction takes at least a few seconds and, more significantly, that some correlations persist over at least several minutes. Notwithstanding, if one takes the mathematical limit where $\tau \to 0$, with $N = T/\tau \to \infty$ but keeping the product $N\sigma_1^2 = T\sigma^2$ finite, one constructs the continuous-time log-normal process, or 'geometrical Brownian motion'. In this limit, using the above form for $P_N(y)$ and $(1+\rho)^N \to e^{rT}$, one

16 In fact, the variance of $P_N(y)$ is equal to $N\sigma_1^2/(1+\rho)^2$, but we neglect this small correction which disappears in the limit $\tau \to 0$.

obtains the celebrated Black–Scholes formula:

$$\mathcal{C}_{\mathrm{BS}}(x_0, x_s, T) = x_0 \int_{y_s}^{\infty} \frac{(\mathrm{e}^{y} - \mathrm{e}^{y_s})}{\sqrt{2\pi\sigma^2 T}} \exp\left(-\frac{(y+\sigma^2 T/2)^2}{2\sigma^2 T}\right) \mathrm{d}y$$

$$= x_0 \mathcal{P}_{G>}\left(\frac{y_-}{\sigma\sqrt{T}}\right) - x_s \mathrm{e}^{-rT} \mathcal{P}_{G>}\left(\frac{y_+}{\sigma\sqrt{T}}\right), \qquad (4.39)$$

where $y_\pm = \log(x_s/x_0) - rT \pm \sigma^2 T/2$ and $\mathcal{P}_{G>}(u)$ is the cumulative normal distribution defined by Eq. (1.68). The way $\mathcal{C}_{\mathrm{BS}}$ varies with the four parameters x_0, x_s, T and σ is quite intuitive, and discussed at length in all the books which deal with options. The derivatives of the price with respect to these parameters are now called the 'Greeks', because they are denoted by Greek letters. For example, the larger the price of the stock x_0, the more probable it is to reach the strike price at expiry, and the more expensive the option. Hence the so-called 'Delta' of the option ($\Delta = \partial\mathcal{C}/\partial x_0$) is positive. As we shall show below, this quantity is actually directly related to the optimal hedging strategy. The variation of Δ with x_0 is noted Γ (Gamma) and is defined as: $\Gamma = \partial\Delta/\partial x_0$. Similarly, the dependence in maturity T and volatility σ (which in fact only appear through the combination $\sigma\sqrt{T}$ if rT is small) leads to the definition of two further 'Greeks': $\Theta = -\partial\mathcal{C}/\partial T = \partial\mathcal{C}/\partial t < 0$ and 'Vega' $\mathcal{V} = \partial\mathcal{C}/\partial\sigma > 0$, the higher $\sigma\sqrt{T}$, the higher the call premium.

Bachelier's Gaussian limit

Suppose now that the price process is additive rather than multiplicative. This means that the increments δx_k themselves, rather than the relative increments, should be considered as independent random variables. Using the change of variable $x_k = (1+\rho)^k \tilde{x}_k$ in Eq. (4.31), one finds that x_N can be written as:

$$x_N = x_0(1+\rho)^N + \sum_{k=0}^{N-1} \delta x_k (1+\rho)^{N-k-1}. \qquad (4.40)$$

When N is large, the difference $x_N - x_0(1+\rho)^N$ becomes, according to the CLT, a Gaussian variable of mean zero (if $m = 0$) and of variance equal to:

$$c_2(T) = D\tau \sum_{\ell=0}^{N-1} (1+\rho)^{2\ell} \simeq DT\left[1 + \rho(N-1) + O(\rho^2 N^2)\right] \qquad (4.41)$$

where $D\tau$ is the variance of the individual increments, related to the volatility through: $D \equiv \sigma^2 x_0^2$. The price, Eq. (4.33), then takes the following form:

$$\mathcal{C}_G(x_0, x_s, T) = \mathrm{e}^{-rT} \int_{x_s}^{\infty} \frac{1}{\sqrt{2\pi c_2(T)}} (x - x_s) \exp\left(-\frac{(x - x_0 \mathrm{e}^{rT})^2}{2c_2(T)}\right) \mathrm{d}x. \quad (4.42)$$

The price formula written down by Bachelier corresponds to the limit of *short maturity* options, where all interest rate effects can be neglected. In this limit, the

above equation simplifies to:

$$\mathcal{C}_G(x_0, x_s, T) \simeq \int_{x_s}^{\infty} (x - x_s) \frac{1}{\sqrt{2\pi DT}} \exp\left(-\frac{(x - x_0)^2}{2DT}\right) \mathrm{d}x. \qquad (4.43)$$

This equation can also be derived directly from the Black–Scholes price, Eq. (4.39), in the small maturity limit, where relative price variations are small: $x_N/x_0 - 1 \ll 1$, allowing one to write $y = \log(x/x_0) \simeq (x - x_0)/x_0$. As emphasized in Section 1.3.2, this is the limit where the Gaussian and log-normal distributions become very similar.

The dependence of $\mathcal{C}_G(x_0, x_s, T)$ as a function of x_s is shown in Figure 4.1, where the numerical value of the relative difference between the Black–Scholes and Bachelier price is also plotted.

In a more general additive model, when N is finite, one can reconstruct the full distribution $P(x, N|x_0, 0)$ from the elementary distribution $P_1(\delta x)$ using the convolution rule (slightly modified to take into account the extra factors $(1+\rho)^{N-k-1}$ in Eq. (4.40)). When N is large, the Gaussian approximation $P(x, N|x_0, 0) = P_G(x, N|x_0(1+\rho)^N, 0)$ becomes accurate, according to the CLT discussed in Chapter 1.

As far as interest rate effects are concerned, it is interesting to note that both formulae, Eqs (4.39) and (4.42), can be written as e^{-rT} times a function of $x_0\mathrm{e}^{rT}$, that is of the forward price $\mathcal{F}_0$. This is quite natural, since an option on a stock and on a forward must have the same price, since they have the same pay-off at expiry. As will be discussed in Chapter 5, this is a rather general property, not related to any particular model for price fluctuations.

Dynamic equation for the option price

It is easy to show directly that the Gaussian distribution:

$$P_G(x, T|x_0, 0) = \frac{1}{\sqrt{2\pi DT}} \exp\left(-\frac{(x - x_0)^2}{2DT}\right), \qquad (4.44)$$

obeys the diffusion equation (or heat equation):

$$\frac{\partial P_G(x, T|x_0, 0)}{\partial T} = \frac{D}{2} \frac{\partial^2 P_G(x, T|x_0, 0)}{\partial x^2}, \qquad (4.45)$$

with boundary conditions:

$$P_G(x, 0|x_0, 0) = \delta(x - x_0). \qquad (4.46)$$

On the other hand, since $P_G(x, T|x_0, 0)$ only depends on the difference $x - x_0$, one has:

$$\frac{\partial^2 P_G(x, T|x_0, 0)}{\partial x^2} = \frac{\partial^2 P_G(x, T|x_0, 0)}{\partial x_0^2}. \qquad (4.47)$$

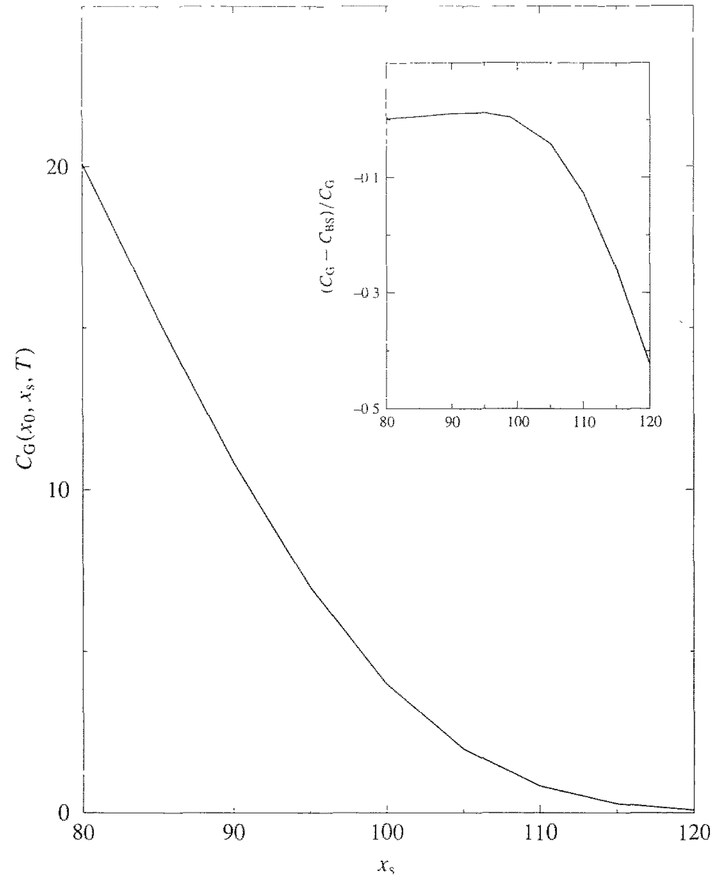

Fig. 4.1. Price of a call of maturity $T = 100$ days as a function of the strike price x_s, in the Bachelier model. The underlying stock is worth 100, and the daily volatility is 1%. The interest rate is taken to be zero. Inset: relative difference between the the log-normal, Black–Scholes price $\mathcal{C}_{BS}$ and the Gaussian, Bachelier price ($\mathcal{C}_G$), for the same values of the parameters.

Taking the derivative of Eq. (4.43) with respect to the maturity T, one finds

$$\frac{\partial \mathcal{C}_G(x_0, x_s, T)}{\partial T} = \frac{D}{2}\frac{\partial^2 \mathcal{C}_G(x_0, x_s, T)}{\partial x_0^2}, \tag{4.48}$$

with boundary conditions, for a zero maturity option:

$$\mathcal{C}_G(x_0, x_s, 0) = \max(x_0 - x_s, 0). \tag{4.49}$$

The option price thus also satisfies the diffusion equation. We shall come back to this point later, in Section 4.5.2: it is indeed essentially this equation that Black and Scholes derived using stochastic calculus in 1973.

4.3.4 Real option prices, volatility smile and 'implied' kurtosis

Stationary distributions and the smile curve

We shall now come back to the case where the distribution of the price increments δx_k is arbitrary, for example a TLD. For simplicity, we set the interest rate ρ to zero. (A non-zero interest rate can readily be taken into account by discounting the price of the call on the forward contract.) The price difference is thus the sum of $N = T/\tau$ iid random variables, to which the discussion of the CLT presented in Chapter 1 applies directly. For large N, the terminal price distribution $P(x, N|x_0, 0)$ becomes Gaussian, while for N finite, 'fat tail' effects are important and the deviations from the CLT are noticeable. In particular, short maturity options or out-of-the-money options, or options on very 'jagged' assets, are not adequately priced by the Black–Scholes formula.

In practice, the market corrects for this difference empirically, by introducing in the Black–Scholes formula an *ad hoc* 'implied' volatility Σ, different from the 'true', historical volatility of the underlying asset. Furthermore, the value of the implied volatility needed to price options of different strike prices x_s and/or maturities T properly is not constant, but rather depends both on T and x_s. This defines a 'volatility surface' $\Sigma(x_s, T)$. It is usually observed that the larger the difference between x_s and x_0, the larger the implied volatility: this is the so-called 'smile effect' (Fig. 4.2). On the other hand, the longer the maturity T, the better the Gaussian approximation; the smile thus tends to flatten out with maturity.

It is important to understand how traders on option markets use the Black–Scholes formula. Rather than viewing it as a predictive theory for option prices given an observed (historical) volatility, the Black–Scholes formula allows the trader to translate back and forth between market prices (driven by supply and demand) and an abstract parameter called the implied volatility. In itself this transformation (between prices and volatilities) does not add any new information. However, this approach is useful in practice, since the implied volatility varies much less than option prices as a function of maturity and moneyness. The Black–Scholes formula is thus viewed as a zero-th order approximation that takes into account the gross features of option prices, the traders then correcting for other effects by adjusting the implied volatility. This view is quite common in experimental sciences where an *a priori* constant parameter of an approximate theory is made into a varying effective parameter to describe more subtle effects.

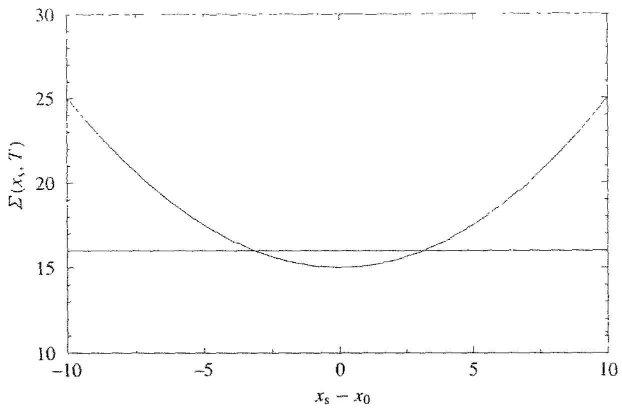

Fig. 4.2. 'Implied volatility' Σ as used by the market to account for non-Gaussian effects. The larger the difference between x_s and x_0, the larger the value of Σ: the curve has the shape of a smile. The function plotted here is a simple parabola, which is obtained theoretically by including the effect of a non-zero kurtosis κ_1 of the elementary price increments. Note that for at-the-money options ($x_s = x_0$), the implied volatility is smaller than the true volatility.

However, the use of the Black–Scholes formula (which assumes a constant volatility) with an *ad hoc* variable volatility is not very satisfactory from a theoretical point of view. Furthermore, this practice requires the manipulation of a whole volatility surface $\Sigma(x_s, T)$, which can deform in time – this is not particularly convenient.

A simple calculation allows one to understand the origin and the shape of the volatility smile. Let us assume that the maturity T is sufficiently large so that only the kurtosis κ_1 of $P_1(\delta x)$ must be taken into account to measure the difference with a Gaussian distribution.[17] Using the results of Section 1.6.3, the formula, Eq. (4.34), leads to:

$$\Delta\mathcal{C}_\kappa(x_0, x_s, T) = \frac{\kappa_1 \tau}{24T}\sqrt{\frac{DT}{2\pi}} \exp\left[-\frac{(x_s - x_0)^2}{2DT}\right]\left(\frac{(x_s - x_0)^2}{DT} - 1\right), \tag{4.50}$$

where $D \equiv \sigma^2 x_0^2$ and $\Delta\mathcal{C}_\kappa = \mathcal{C}_\kappa - \mathcal{C}_{\kappa=0}$.

One can indeed transform the formula

$$\mathcal{C} = \int_{x_s}^{\infty} (x' - x_s) P(x', T|x_0, 0)\,\mathrm{d}x', \tag{4.51}$$

[17] We also assume here that the distribution is symmetrical ($P_1(\delta x) = P_1(-\delta x)$), which is usually justified on short time scales. If this assumption is not adequate, one must include the skewness λ_3, leading to an asymmetrical smile, cf. Eq. (4.56).

through an integration by parts

$$\mathcal{C} = \int_{x_s}^{\infty} \mathcal{P}_>(x', T|x_0, 0)\,\mathrm{d}x'. \tag{4.52}$$

After changing variables $x' \to x' - x_0$, and using Eq. (1.69) of Section 1.6.3, one gets:

$$\begin{aligned}\mathcal{C} &= \mathcal{C}_G + \frac{\sqrt{DT}}{\sqrt{N}} \int_{u_s}^{\infty} \frac{1}{\sqrt{2\pi}} Q_1(u) \mathrm{e}^{-u^2/2}\,\mathrm{d}u \\ &\quad + \frac{\sqrt{DT}}{N} \int_{u_s}^{\infty} \frac{1}{\sqrt{2\pi}} Q_2(u) \mathrm{e}^{-u^2/2}\,\mathrm{d}u + \cdots\end{aligned} \tag{4.53}$$

where $u_s \equiv (x_s - x_0)/\sqrt{DT}$. Now, noticing that:

$$Q_1(u)\mathrm{e}^{-u^2/2} = \frac{\lambda_3}{6}\frac{\mathrm{d}^2}{\mathrm{d}u^2}\mathrm{e}^{-u^2/2}, \tag{4.54}$$

and

$$Q_2(u)\mathrm{e}^{-u^2/2} = -\frac{\lambda_4}{24}\frac{\mathrm{d}^3}{\mathrm{d}u^3}\mathrm{e}^{-u^2/2} - \frac{\lambda_3^2}{72}\frac{\mathrm{d}^5}{\mathrm{d}u^5}\mathrm{e}^{-u^2/2}, \tag{4.55}$$

the integrations over u are readily performed, yielding:

$$\begin{aligned}\mathcal{C} &= \mathcal{C}_G + \sqrt{DT}\frac{\mathrm{e}^{-u_s^2/2}}{\sqrt{2\pi}}\left(\frac{\lambda_3}{6\sqrt{N}}u_s + \frac{\lambda_4}{24N}(u_s^2 - 1)\right.\\ &\quad \left. + \frac{\lambda_3^2}{72N}(u_s^4 - 6u_s^2 + 3) + \cdots\right),\end{aligned} \tag{4.56}$$

which indeed coincides with Eq. (4.50) for $\lambda_3 = 0$. In general, one has $\lambda_3^2 \ll \lambda_4$; in this case, the smile remains a parabola, but shifted and centred around $x_0(1 - 2\sigma T\lambda_3/\lambda_4)$.

Note that we refer here to *additive* (rather than multiplicative) price increments: the Black–Scholes volatility smile is often asymmetrical merely because of the use of a skewed log-normal distribution to describe a nearly symmetrical process.

On the other hand, a simple calculation shows that the variation of $\mathcal{C}_{\kappa=0}(x_0, x_s, T)$ [as given by Eq. (4.43)] when the volatility changes by a small quantity $\delta D = 2\sigma x_0^2 \delta\sigma$ is given by:

$$\delta\mathcal{C}_{\kappa=0}(x_0, x_s, T) = \delta\sigma x_0 \sqrt{\frac{T}{2\pi}} \exp\left[-\frac{(x_s - x_0)^2}{2DT}\right]. \tag{4.57}$$

The effect of a non-zero kurtosis κ_1 can thus be reproduced (to first order) by a Gaussian pricing formula, but at the expense of using an effective volatility $\Sigma(x_s, T) = \sigma + \delta\sigma$ given by:

$$\Sigma(x_s, T) = \sigma\left[1 + \frac{\kappa(T)}{24}\left(\frac{(x_s - x_0)^2}{DT} - 1\right)\right], \tag{4.58}$$

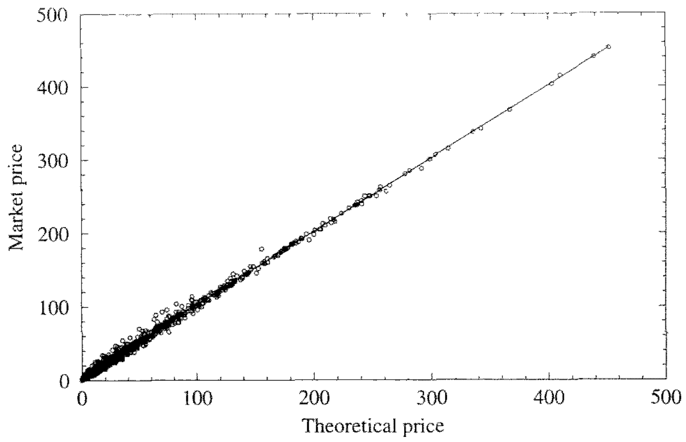

Fig. 4.3. Comparison between theoretical and observed option prices. Each point corresponds to an option on the Bund (traded on the LIFFE in London), with different strike prices and maturities. The x coordinate of each point is the theoretical price of the option (determined using the empirical terminal price distribution), while the y coordinate is the market price of the option. If the theory is good, one should observe a cloud of points concentrated around the line $y = x$. The dotted line corresponds to a linear regression, and gives $y = 0.998x + 0.02$ (in basis points units).

with $\kappa(T) = \kappa_1/N$. This very simple formula, represented in Figure 4.2, allows one to understand intuitively the shape and amplitude of the smile. For example, for a daily kurtosis of $\kappa_1 = 10$, the volatility correction is on the order of $\delta\sigma/\sigma \simeq 17\%$ for out-of-the-money options such that $x_s - x_0 = 3\sqrt{DT}$, and for $T = 20$ days. Note that the effect of the kurtosis is to *reduce* the implied at-the-money volatility in comparison with its true value.

Figure 4.3 shows some 'experimental' data, concerning options on the Bund (futures) contract – for which a weakly non-Gaussian model is satisfactory (cf. Section 2.3). The associated option market is, furthermore, very liquid; this tends to reduce the width of the bid-ask spread, and thus, probably to decrease the difference between the market price and a theoretical 'fair' price. This is not the case for OTC options, when the overhead charged by the financial institution writing the option can in some case be very large, cf. Section 4.4.1 below. In Figure 4.3, each point corresponds to an option of a certain maturity and strike price. The coordinates of these points are the theoretical price, Eq. (4.34), along the x axis (calculated using the historical distribution of the Bund), and the observed market price along the y axis. If the theory is good, one should observe a cloud of points concentrated around the line $y = x$. Figure 4.3 includes points from the first half of 1995, which was rather 'calm', in the sense that the volatility remained roughly constant during that period (see Fig. 4.4). Correspondingly, the assumption that the process is stationary is reasonable.

The agreement between theoretical and observed prices is however much less convincing if one uses data from 1994, when the volatility of interest rate markets has been high. The following subsection aims at describing these discrepancies in greater details.

Non-stationarity and 'implied' kurtosis

A more precise analysis reveals that the scale of the fluctuations of the underlying asset (here the Bund contract) itself varies noticeably around its mean value; these 'scale fluctuations' are furthermore correlated with the option prices. More precisely, one postulates that the distribution of price increments δx has a constant shape, but a width γ_k which is time dependent (cf. Section 2.4), i.e.:[18]

$$P_1(\delta x_k) \equiv \frac{1}{\gamma_k} P_{10}\left(\frac{\delta x_k}{\gamma_k}\right), \tag{4.59}$$

where P_{10} is a distribution independent of k, and of width (for example measured by the MAD) normalized to one. The absolute volatility of the asset is thus proportional to γ_k. Note that if P_{10} is Gaussian and time is continuous, this model is known as the stochastic volatility Brownian motion.[19] *However, this assumption is not needed, and one can keep P_{10} arbitrary.*

Figure 4.4 shows the evolution of the scale γ (filtered over 5 days in the past) as a function of time and compares it to the implied volatility $\Sigma(x_s = x_0)$ extracted from at-the-money options during the same period. One thus observes that the option prices are rather well tracked by adjusting the factor γ_k through a short-term estimate of the volatility of the underlying asset. This means that the option prices primarily reflects the quasi-instantaneous volatility of the underlying asset itself, rather than an 'anticipated' average volatility on the life time of the option.

It is interesting to notice that the mere existence of volatility fluctuations leads to a non-zero kurtosis κ_N of the asset fluctuations (see Section 2.4). This kurtosis has an anomalous time dependence (cf. Eq. (2.17)), in agreement with the direct observation of the historical kurtosis (Fig. 4.6). On the other hand, an 'implied kurtosis' can be extracted from the market price of options, using Eq. (4.58) as a fit to the empirical smile, see Figure 4.5. Remarkably enough, this implied kurtosis is in close correspondence with the historical kurtosis – note that Figure 4.6 does not require any further adjustable parameter.

As a conclusion of this section, it seems that market operators have empirically corrected the Black–Scholes formula to account for two distinct, but related effects:

- *The presence of market jumps, implying fat tailed distributions ($\kappa > 0$) of short-term price increments. This effect is responsible for the volatility smile (and also, as we shall discuss next, for the fact that options are risky).*
- *The fact that the volatility is not constant, but fluctuates in time, leading to an anomalously slow decay of the kurtosis (slower than $1/N$) and, correspondingly, to a non-trivial deformation of the smile with maturity.*

18 This hypothesis can be justified by assuming that the amplitude of the market moves is subordinated to the volume of transactions, which itself is obviously time dependent.

19 In this context, the fact that the volatility is time dependent is called 'heteroskedasticity'. ARCH models (Auto Regressive Conditional Heteroskedasticity) and their relatives have been invented as a framework to model such effects. See Section 2.9.

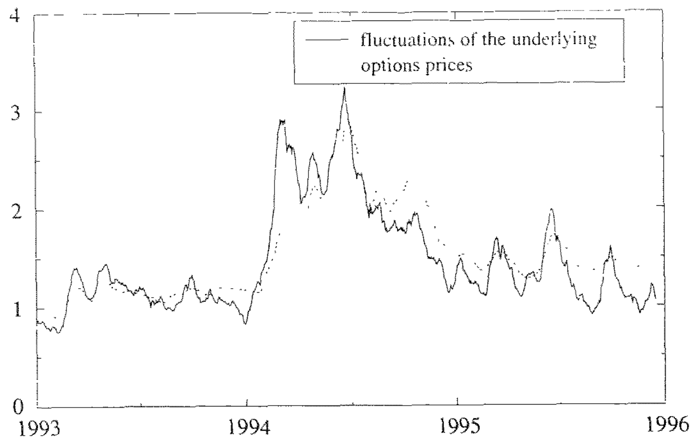

Fig. 4.4. Time dependence of the scale parameter γ, obtained from the analysis of the intra-day fluctuations of the underlying asset (here the Bund), or from the implied volatility of at-the-money options. More precisely, the historical determination of γ comes from the daily average of the absolute value of the 5-min price increments. The two curves are then smoothed over 5 days. These two determinations are strongly correlated, showing that the option price mostly reflects the instantaneous volatility of the underlying asset itself.

It is interesting to note that, through trial and errors, the market as a whole has evolved to allow for such non-trivial statistical features – at least on most actively traded markets. This might be called 'market efficiency'; but contrarily to stock markets where it is difficult to judge whether the stock price is or is not the 'true' price (which might be an empty concept), option markets offer a remarkable testing ground for this idea. It is also a nice example of adaptation of a population (the traders) to a complex and hostile environment, which has taken place in a few decades!

In summary, part of the information contained in the implied volatility surface $\Sigma(x_s, T)$ used by market participants can be explained by an adequate statistical model of the underlying asset fluctuations. In particular, in weakly non-Gaussian markets, an important parameter is the time-dependent kurtosis, see Eq. (4.58). The anomalous maturity dependence of this kurtosis encodes the fact that the volatility is itself time dependent.

4.4 Optimal strategy and residual risk

4.4.1 Introduction

In the above discussion, we have chosen a model of price increments such that the cost of the hedging strategy (i.e. the term $\langle \psi_k \delta x_k \rangle$) could be neglected, which is justified if the excess return m is zero, or else for short maturity options. Beside the

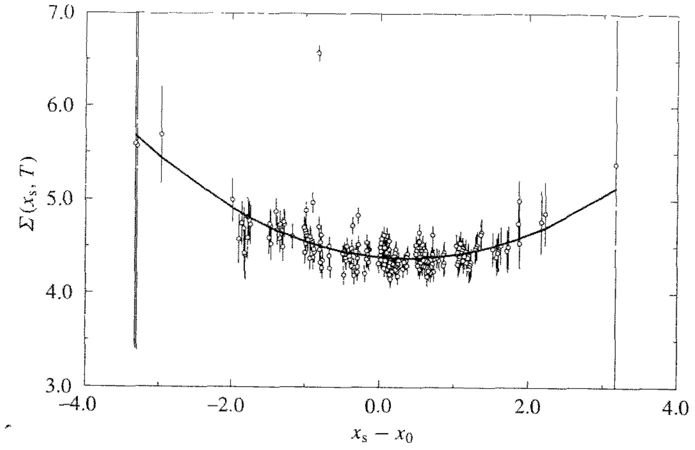

Fig. 4.5. Implied Black–Scholes volatility and fitted parabolic smile. The circles correspond to all quoted prices on 26th April, 1995, on options of 1-month maturity. The error bars correspond to an error on the price of ± 1 basis point. The curvature of the parabola allows one to extract the implied kurtosis $\kappa(T)$ using Eq. (4.58).

fact that the correction to the option price induced by a non-zero return is important to assess (this will be done in the next section), the determination of an 'optimal' hedging strategy and the corresponding minimal residual risk is crucial for the following reason. Suppose that an adequate measure of the risk taken by the writer of the option is given by the variance of the global wealth balance associated to the operation, i.e.:[20]

$$\mathcal{R} = \sqrt{\langle \Delta W^2[\phi] \rangle}. \tag{4.60}$$

As we shall find below, there is a special strategy ϕ^* such that the above quantity is reaches a minimum value. Within the hypotheses of the Black–Scholes model, this minimum risk is even, rather surprisingly, *strictly zero*. Under less restrictive and more realistic hypotheses, however, the residual risk $\mathcal{R}^* \equiv \sqrt{\langle \Delta W^2[\phi^*] \rangle}$ actually amounts to a substantial fraction of the option price itself. It is thus rather natural for the writer of the option to try to reduce further the risk by overpricing the option, adding to the 'fair price' a risk premium proportional to $\mathcal{R}^*$ – in such a way that the probability of eventually losing money is reduced. Stated otherwise, option writing being an essentially risky operation, it should also be, on average, profitable.

Therefore, a market maker on option markets will offer to buy and to sell options at slightly different prices (the 'bid–ask' spread), centred around the fair price $\mathcal{C}$. The amplitude of the spread is presumably governed by the residual risk, and is

[20] The case where a better measure of risk is the loss probability or the value-at-risk will be discussed below.

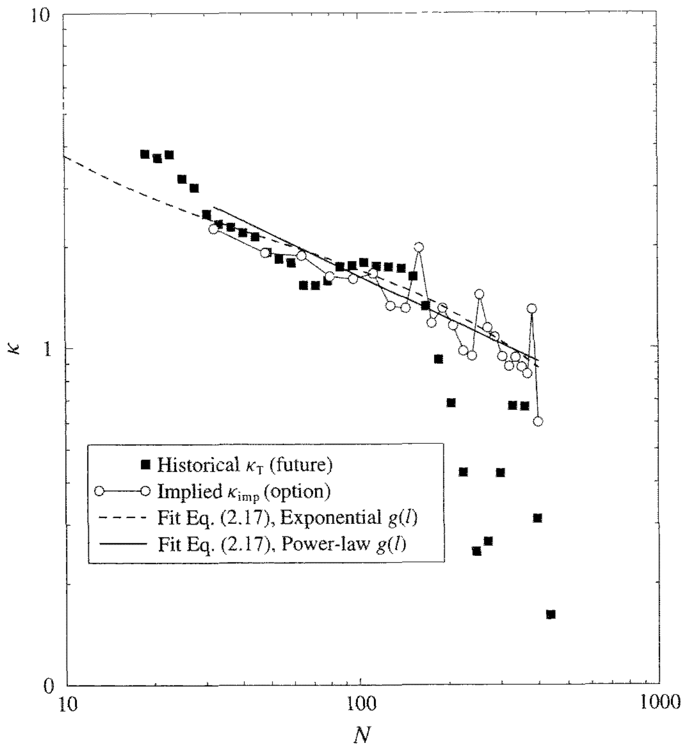

Fig. 4.6. Plot (in log–log coordinates) of the average implied kurtosis κ_{imp} (determined by fitting the implied volatility for a fixed maturity by a parabola) and of the empirical kurtosis κ_N (determined directly from the historical movements of the Bund contract), as a function of the reduced time scale $N = T/\tau$, $\tau = 30$ min. All transactions of options on the Bund future from 1993 to 1995 were analysed along with 5 minute tick data of the Bund future for the same period. We show for comparison a fit with $\kappa_N \simeq N^{-0.42}$ (dark line), as suggested by the results of Section 2.4. A fit with an exponentially decaying volatility correlation function is however also acceptable (dash line).

thus $\pm\lambda\mathcal{R}^*$, where λ is a certain numerical factor, which measures the price of risk. The search for minimum risk corresponds to the need of keeping the bid–ask spread as small as possible, because several competing market makers are present. One therefore expects the coefficient λ to be smaller on liquid markets. On the contrary, the writer of an OTC option usually claims a rather high risk premium λ.

Let us illustrate this idea by Figures 4.7 and 4.8, generated using real market data. We show the histogram of the global wealth balance ΔW, corresponding to an at-the-money option, of maturity equal to 3 months, in the case of a bare

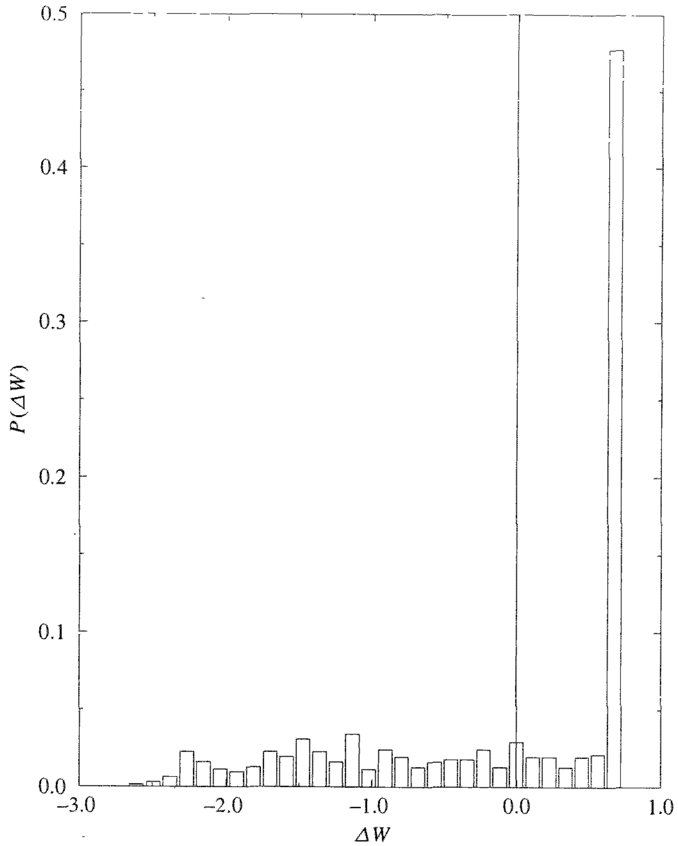

Fig. 4.7. Histogram of the global wealth balance ΔW associated with the writing of an at-the-money option of maturity equal to 60 trading days. The price is fixed such that on average $\langle \Delta W \rangle$ is zero (vertical line). This figure corresponds to the case where the option is not hedged ($\phi \equiv 0$). The RMS of the distribution is 1.04, to be compared with the price of the option $\mathcal{C} = 0.79$. The 'peak' at 0.79 thus corresponds to the cases where the option is not exercised, which happens with a probability close to $\frac{1}{2}$ for at-the-money options.

position ($\phi \equiv 0$), and in the case where one follows the optimal strategy ($\phi = \phi^*$) prescribed below. The fair price of the option is such that $\langle \Delta W \rangle = 0$ (vertical thick line). It is clear that without hedging, option writing is a very risky operation. The optimal hedge substantially reduces the risk, though the residual risk remains quite high. Increasing the price of the option by an amount $\lambda\mathcal{R}^*$ corresponds to a shift of the vertical thick line to the left of Figure 4.8, thereby reducing the weight of unfavourable events.

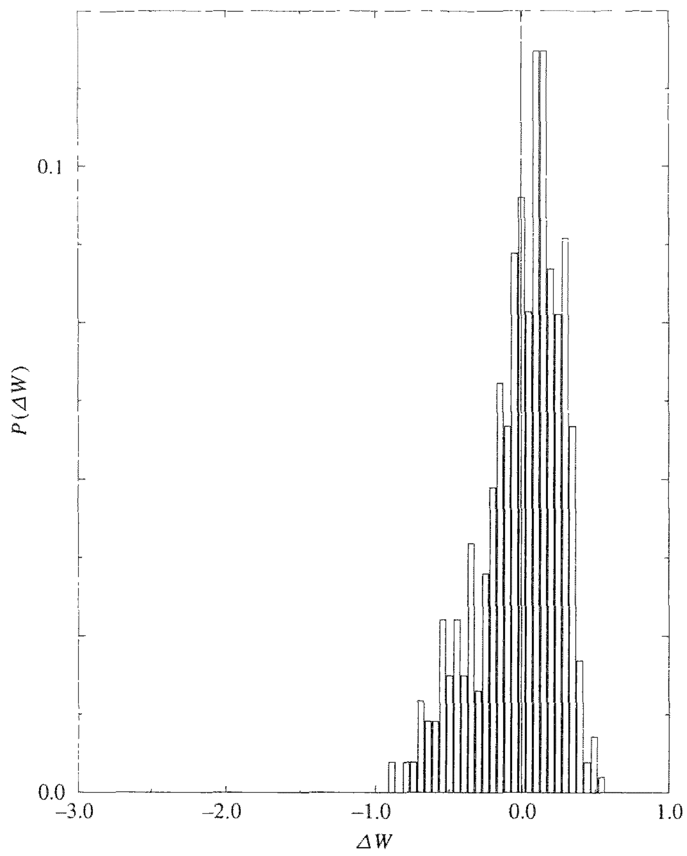

Fig. 4.8. Histogram of the global wealth balance ΔW associated to the writing of the same option, with a price still fixed such that $\langle \Delta W \rangle = 0$ (vertical line), with the same horizontal scale as in the previous figure. This figure shows the effect of adopting an optimal hedge ($\phi = \phi^*$), recalculated every half-hour, and in the absence of transaction costs. The RMS $\mathcal{R}^*$ is clearly smaller ($= 0.28$), but non-zero. Note that the distribution is skewed towards $\Delta W < 0$. Increasing the price of the option by $\lambda \mathcal{R}^*$ amounts to diminishing the probability of losing money (for the writer of the option), $\mathcal{P}(\Delta W < 0)$.

Another way of expressing this idea is to use $\mathcal{R}^*$ as a scale to measure the difference between the market price of an option $\mathcal{C}_M$ and its theoretical price:

$$\lambda = \frac{\mathcal{C}_M - \mathcal{C}}{\mathcal{R}^*}. \tag{4.61}$$

An option with a large value of λ is an expensive option, which includes a large risk premium.

4.4.2 A simple case

Let us now discuss how an optimal hedging strategy can be constructed, by focusing first on the simple case where the amount of the underlying asset held in the portfolio is fixed once and for all when the option is written, i.e. at $t = 0$. This extreme case would correspond to very high transaction costs, so that changing one's position on the market is very costly. We shall furthermore assume, for simplicity, that interest rate effects are negligible, i.e. $\rho = 0$. The global wealth balance, Eq. (4.28), then reads:

$$\Delta W = \mathcal{C} - \max(x_N - x_s, 0) + \phi \sum_{k=0}^{N-1} \delta x_k. \tag{4.62}$$

In the case where the average return is zero ($\langle \delta x_k \rangle = 0$) and where the increments are uncorrelated (i.e. $\langle \delta x_k \delta x_l \rangle = D\tau\delta_{k,l}$), the variance of the final wealth, $\mathcal{R}^2 = \langle \Delta W^2 \rangle - \langle \Delta W \rangle^2$, reads:

$$\mathcal{R}^2 = N D\tau\phi^2 - 2\phi\langle (x_N - x_0)\max(x_N - x_s, 0)\rangle + \mathcal{R}_0^2, \tag{4.63}$$

where $\mathcal{R}_0^2$ is the intrinsic risk, associated to the 'bare', unhedged option ($\phi = 0$):

$$\mathcal{R}_0^2 = \langle \max(x_N - x_s, 0)^2 \rangle - \langle \max(x_N - x_s, 0)\rangle^2. \tag{4.64}$$

The dependence of the risk on ϕ is shown in Figure 4.9, and indeed reveals the existence of an optimal value $\phi = \phi^*$ for which $\mathcal{R}$ takes a minimum value. Taking the derivative of Eq. (4.63) with respect to ϕ,

$$\left.\frac{d\mathcal{R}}{d\phi}\right|_{\phi=\phi^*} = 0, \tag{4.65}$$

one finds:

$$\phi^* = \frac{1}{D\tau N} \int_{x_s}^{\infty} (x - x_s)(x - x_0) P(x, N|x_0, 0)\, dx, \tag{4.66}$$

thereby fixing the optimal hedging strategy within this simplified framework. Interestingly, if $P(x, N|x_0, 0)$ is Gaussian (which becomes a better approximation as $N = T/\tau$ increases), one can write:

$$\frac{1}{DT}\int_{x_s}^{\infty} \frac{1}{\sqrt{2\pi DT}}(x - x_s)(x - x_0)\exp\left[-\frac{(x - x_0)^2}{2DT}\right] dx \tag{4.67}$$

$$= -\int_{x_s}^{\infty} \frac{1}{\sqrt{2\pi DT}}(x - x_s)\frac{\partial}{\partial x}\exp\left[-\frac{(x - x_0)^2}{2DT}\right] dx,$$

giving, after an integration by parts: $\phi^* \equiv \mathcal{P}$, or else the probability, calculated from $t = 0$, that the option is exercised at maturity: $\mathcal{P} \equiv \int_{x_s}^{\infty} P(x, N|x_0, 0)\, dx$.

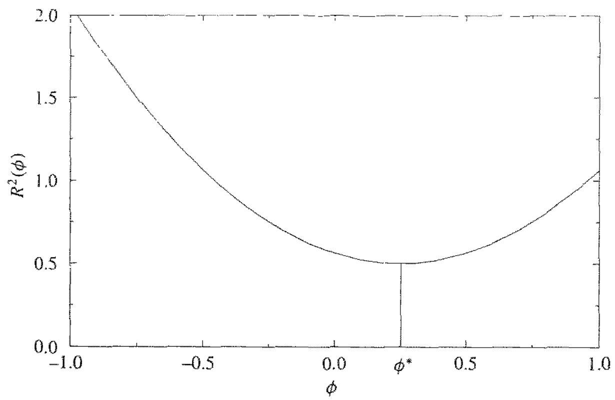

Fig. 4.9. Dependence of the residual risk $\mathcal{R}$ as a function of the strategy ϕ, in the simple case where this strategy does not vary with time. Note that $\mathcal{R}$ is minimum for a well-defined value of ϕ.

Hence, buying a certain fraction of the underlying allows one to reduce the risk associated to the option. As we have formulated it, risk minimization appears as a *variational problem.* The result therefore depends on the family of 'trial' strategies ϕ. A natural idea is thus to generalize the above procedure to the case where the strategy ϕ is allowed to vary both with time, and with the price of the underlying asset. In other words, one can certainly do better than holding a certain fixed quantity of the underlying asset, by adequately readjusting this quantity in the course of time.

4.4.3 General case: 'Δ' hedging

If one writes again the complete wealth balance, Eq. (4.26), as:

$$\Delta W = \mathcal{C}(1+\rho)^N - \max(x_N - x_s, 0) + \sum_{k=0}^{N-1} \psi_k^N(x_k)\delta x_k, \tag{4.68}$$

the calculation of $\langle \Delta W^2 \rangle$ involves, as in the simple case treated above, three types of terms: quadratic in ψ, linear in ψ, and independent of ψ. The last family of terms can thus be forgotten in the minimization procedure. The first two terms of $\mathcal{R}^2$ are:

$$\sum_{k=0}^{N-1} \langle (\psi_k^N)^2 \rangle \langle \delta x_k^2 \rangle - 2 \sum_{k=0}^{N-1} \langle \psi_k^N \delta x_k \max(x_N - x_s, 0) \rangle, \tag{4.69}$$

where we have used $\langle \delta x_k \rangle = 0$ and assumed the increments to be of finite variance and uncorrelated (but not necessarily stationary nor independent):[21]

$$\langle \delta x_k \delta x_l \rangle = \langle \delta x_k^2 \rangle \delta_{k,l}. \tag{4.70}$$

We introduce again the distribution $P(x, k|x_0, 0)$ for arbitrary intermediate times k. The strategy ψ_k^N depends now on the value of the price x_k. One can express the terms appearing in Eq. (4.69) in the following form:

$$\left\langle (\psi_k^N)^2 \right\rangle \langle \delta x_k^2 \rangle = \int [\psi_k^N(x)]^2 P(x, k|x_0, 0) \langle \delta x_k^2 \rangle \, \mathrm{d}x, \tag{4.71}$$

and

$$\left\langle \psi_k^N \delta x_k \max(x_N - x_s, 0) \right\rangle = \int \psi_k^N(x) P(x, k|x_0, 0) \, \mathrm{d}x \times \int_{x_s}^{+\infty} \langle \delta x_k \rangle_{(x,k)\to(x',N)} (x' - x_s) P(x', N|x, k) \, \mathrm{d}x', \tag{4.72}$$

where the notation $\langle \delta x_k \rangle_{(x,k)\to(x',N)}$ means that the average of δx_k is restricted to those trajectories which start from point x at time k and end at point x' at time N. Without this restriction, the average would of course be (within the present hypothesis) zero.

The *functions* $\psi_k^N(x)$, for $k = 1, 2, \dots, N$, must be chosen so that the risk $\mathcal{R}$ is as small as possible. Technically, one must 'functionally' minimize Eq. (4.69). We shall not try to justify this procedure mathematically; one can simply imagine that the integrals over x are discrete sums (which is actually true, since prices are expressed in cents and therefore are not continuous variables). The function $\psi_k^N(x)$ is thus determined by the discrete set of values of $\psi_k^N(i)$. One can then take usual derivatives of Eq. (4.69) with respect to these $\psi_k^N(i)$. The continuous limit, where the points i become infinitely close to one another, allows one to define the functional derivative $\partial/\partial\psi_k^N(x)$, but it is useful to keep in mind its discrete interpretation. Using Eqs. (4.71) and (4.73), one thus finds the following fundamental equation:

$$\frac{\partial \mathcal{R}}{\partial \psi_k^N(x)} = 2\psi_k^N(x) P(x, k|x_0, 0) \langle \delta x_k^2 \rangle - 2P(x, k|x_0, 0) \int_{x_s}^{+\infty} \langle \delta x_k \rangle_{(x,k)\to(x',N)} (x' - x_s) P(x', N|x, k) \, \mathrm{d}x'. \tag{4.73}$$

Setting this functional derivative to zero then provides a general and rather explicit expression for the optimal strategy $\psi_k^{N*}(x)$, where the *only assumption* is that the

[21] The case of correlated Gaussian increments is considered in [Bouchaud and Sornette].

increments δx_k are uncorrelated (cf. Eq. (4.70)):[22]

$$\psi_k^{N*}(x) = \frac{1}{\langle \delta x_k^2 \rangle} \int_{x_s}^{+\infty} \langle \delta x_k \rangle_{(x,k)\to(x',N)} (x' - x_s) P(x', N|x, k)\, dx'. \tag{4.74}$$

This formula can be simplified in the case where the increments δx_k are identically distributed (of variance $D\tau$) and when the interest rate is zero. As shown in Appendix D, one then has:

$$\langle \delta x_k \rangle_{(x,k)\to(x',N)} = \frac{x' - x}{N - k}. \tag{4.75}$$

The intuitive interpretation of this result is that all the $N - k$ increments δx_l along the trajectory $(x, k) \to (x', N)$ contribute on average equally to the overall price change $x' - x$. The optimal strategy then finally reads:

$$\phi_k^{N*}(x) = \int_{x_s}^{+\infty} \frac{x' - x}{D\tau(N-k)} (x' - x_s) P(x', N|x, k)\, dx'. \tag{4.76}$$

We leave to the reader to show that the above expression varies monotonically from $\phi_k^{N*}(-\infty) = 0$ to $\phi_k^{N*}(+\infty) = 1$; this result holds without any further assumption on $P(x', N|x, k)$. If $P(x', N|x, k)$ is well approximated by a Gaussian, the following identity then holds true:

$$\frac{x' - x}{D\tau(N-k)} P_G(x', N|x, k) \equiv \frac{\partial P_G(x', N|x, k)}{\partial x}. \tag{4.77}$$

The comparison between Eqs (4.76) and (4.33) then leads to the so-called 'Delta' hedging of Black and Scholes:

$$\begin{aligned} \phi_k^{N*}(x = x_k) &= \left.\frac{\partial \mathcal{C}_{BS}[x, x_s, N-k]}{\partial x}\right|_{x=x_k} \\ &\equiv \Delta(x_k, N-k). \end{aligned} \tag{4.78}$$

One can actually show that this result is true even if the interest rate is non-zero (see Appendix D), as well as if the relative returns (rather than the absolute returns) are independent Gaussian variables, as assumed in the Black–Scholes model.

Equation (4.78) has a very simple (perhaps sometimes misleading) interpretation. If between time k and $k+1$ the price of the underlying asset varies by a small amount dx_k, then to first order in dx_k, the variation of the option price is given by $\Delta[x, x_s, N-k]\, dx_k$ (using the very definition of Δ). This variation is therefore exactly compensated by the gain or loss of the strategy, i.e. $\phi_k^{N*}(x = x_k)\, dx_k$. In other words, $\phi_k^{N*} = \Delta(x_k, N-k)$ appears to be a perfect hedge, which ensures that

[22] This is true within the variational space considered here, where ϕ only depends on x and t. If some volatility correlations are present, one could in principle consider strategies that depend on the past values of the volatility.

the portfolio of the writer of the option does not vary *at all* with time (cf. below, when we will discuss the differential approach of Black and Scholes). However, as we discuss now, the relation between the optimal hedge and Δ is not general, and does not hold for non-Gaussian statistics.

Cumulant corrections to Δ hedging

More generally, using Fourier transforms, one can express Eq. (4.76) as a cumulant expansion. Using the definition of the cumulants of P, one has:

$$\begin{aligned} (x' - x)P(x', N|x, 0) &\equiv \frac{1}{2\pi} \int \hat{P}_N(z) \frac{\partial}{\partial(-iz)} e^{-iz(x'-x)}\, dz \\ &= -\sum_{n=2}^{\infty} \frac{(-)^n c_{n,N}}{(n-1)!} \frac{\partial^{n-1}}{\partial x'^{n-1}} P(x', N|x, 0), \end{aligned} \tag{4.79}$$

where $\log \hat{P}_N(z) = \sum_{n=2}^{\infty} (iz)^n c_{n,N}/n!$. Assuming that the increments are independent, one can use the additivity property of the cumulants discussed in Chapter 1, i.e.: $c_{n,N} \equiv N c_{n,1}$, where the $c_{n,1}$ are the cumulants at the elementary time scale τ. One finds the following general formula for the optimal strategy:[23]

$$\phi^{N*}(x) = \frac{1}{D\tau} \sum_{n=2}^{\infty} \frac{c_{n,1}}{(n-1)!} \frac{\partial^{n-1} \mathcal{C}[x, x_s, N]}{\partial x^{n-1}}, \tag{4.80}$$

where $c_{2,1} \equiv D\tau$, $c_{4,1} \equiv \kappa_1 [c_{2,1}]^2$, etc. In the case of a Gaussian distribution, $c_{n,1} \equiv 0$ for all $n \geq 3$, and one thus recovers the previous 'Delta' hedging. If the distribution of the elementary increments δx_k is symmetrical, $c_{3,1} = 0$. The first correction is then of order $c_{4,1}/c_{2,1} \times (1/\sqrt{D\tau N})^2 \equiv \kappa/N$, where we have used the fact that $\mathcal{C}[x, x_s, N]$ typically varies on the scale $x \simeq \sqrt{D\tau N}$, which allows us to estimate the order of magnitude of its derivatives with respect to x. Equation (4.80) shows that in general, the optimal strategy is not simply given by the derivative of the price of the option with respect to the price of the underlying asset.

It is interesting to compute the difference between ϕ^* and the Black–Scholes strategy as used by the traders, ϕ_M^*, which takes into account the implied volatility of the option $\Sigma(x_s, T = N\tau)$:[24]

$$\phi_M^*(x, \Sigma) \equiv \left.\frac{\partial \mathcal{C}[x, x_s, T]}{\partial x}\right|_{\sigma=\Sigma}. \tag{4.81}$$

If one chooses the implied volatility Σ in such a way that the 'correct' price (cf. Eq. (4.58) above) is reproduced, one can show that:

$$\phi^*(x) = \phi_M^*(x, \Sigma) + \frac{\kappa_1 \tau}{12T} \frac{(x_s - x_0)}{\sqrt{2\pi DT}} \exp\left[-\frac{(x_s - x_0)^2}{2DT}\right], \tag{4.82}$$

where higher-order cumulants are neglected. The Black–Scholes strategy, even calculated

[23] This formula can actually be generalized to the case where the volatility is time dependent, see Appendix D and Section 5.2.3.

[24] Note that the market does not compute the total derivative of the option price with respect to the underlying price, which would also include the derivative of the implied volatility.

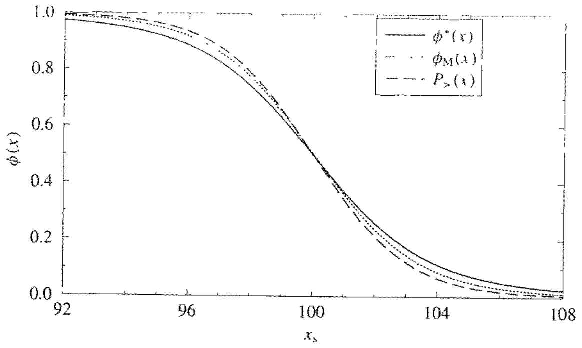

Fig. 4.10. Three hedging strategies for an option on an asset of terminal value x_N distributed according to a truncated Lévy distribution, of kurtosis 5, as a function of the strike price x_s. ϕ^* denotes the optimal strategy, calculated from Eq. (4.76), ϕ_M is the Black–Scholes strategy computed with the appropriate implied volatility, such as to reproduce the correct option price, and $\mathcal{P}_>$ is the probability that the option is exercised at expiry, which would be the optimal strategy if the fluctuations were Gaussian. These three strategies are actually quite close to each other (they coincide at-the-money, and deep in-the-money and out-the-money). Note however that, interestingly, the variations of ϕ^* with the price are slower (smaller Γ).

with the implied volatility, is thus not the optimal strategy when the kurtosis is non-zero. However, the difference between the two reaches a maximum when $x_s - x_0 = \sqrt{DT}$, and is equal to:

$$\delta\phi^* \simeq 0.02\frac{\kappa_1}{N} \qquad T = N\tau. \tag{4.83}$$

For a daily kurtosis of $\kappa_1 = 10$ and for $N = 20$ days, one finds that the first correction to the optimal hedge is on the order of 0.01, see also Figure 4.10. As will be discussed below, the use of such an approximate hedging induces a rather small increase of the residual risk.

4.4.4 Global hedging/instantaneous hedging

It is important to stress that the above procedure, where the global risk associated to the option (calculated as the variance of the final wealth balance) is minimized is *equivalent* to the minimization of an 'instantaneous' hedging error – at least when the increments are uncorrelated. The latter consists in minimizing the difference of value of a position which is short one option and ϕ long in the underlying stock, between two consecutive times $k\tau$ and $(k+1)\tau$. This is closer to the concern of many option traders, since the value of their option books is calculated every day: the risk is estimated in a 'marked to market' way. If the price increments are

uncorrelated, we show below that the optimal strategy is indeed identical to the 'global' one discussed above.

The change of wealth between times k and k + 1 is, in the absence of interest rates, given by:

$$\delta W_k = \mathcal{C}_k - \mathcal{C}_{k+1} + \phi_k(x_k)\delta x_k. \tag{4.84}$$

The global wealth balance is simply given by the sum over n of all these δW_k. Let us now calculate the part of $\langle \delta W_k^2 \rangle$ which depends on ϕ_k. Using the fact that the increments are uncorrelated, one finds:

$$\phi_k^2(x_k)D\tau - 2\phi_k(x_k)\langle \mathcal{C}_{k+1}\delta x_k \rangle. \tag{4.85}$$

Using now the explicit expression for $\mathcal{C}_{k+1}$:

$$\mathcal{C}_{k+1} = \int_{x_s}^{\infty} (x' - x_s)P(x', N|x_{k+1}, k+1)\,\mathrm{d}x', \tag{4.86}$$

one sees that the second term in the above expression contains the following average:

$$\int (x_{k+1} - x_k)P(x_{k+1}, k+1|x_k, k)P(x', N|x_{k+1}, k+1)\,\mathrm{d}x_{k+1}. \tag{4.87}$$

Using the methods of Appendix D, one can show that the above average is equal to $(x' - x_k/N - k)P(x', N|x_k, k)$. Therefore, the part of the risk that depends on ϕ_k is given by:

$$\phi_k^2(x_k)D\tau - 2\phi_k(x_k)\int_{x_s}^{\infty} \frac{(x' - x_s)(x' - x_k)}{N - k} P(x', N|x_k, k)\,\mathrm{d}x'. \tag{4.88}$$

Taking the derivative of this expression with respect to ϕ_k finally leads to the optimal strategy, Eq. (4.76), above.

4.4.5 Residual risk: the Black–Scholes miracle

We shall now compute the *residual risk* $\mathcal{R}^*$ obtained by substituting ϕ by ϕ^* in Eq. (4.69). One finds:

$$\mathcal{R}^{*2} = \mathcal{R}_0^2 - D\tau \sum_{k=0}^{N-1} \int P(x, k|x_0, 0)[\phi_k^{N*}(x)]^2\,\mathrm{d}x, \tag{4.89}$$

where $\mathcal{R}_0$ is the unhedged risk. The Black–Scholes 'miracle' is the following: in the case where P is Gaussian, the two terms in the above equation *exactly compensate* in the limit where $\tau \to 0$, with $N\tau = T$ fixed.[25]

[25] The 'zero-risk' property is true both within an 'additive' Gaussian model and the Black–Scholes multiplicative model, or for more general continuous-time Brownian motions. This property is more easily obtained using Ito's stochastic differential calculus, see below. The latter formalism is however only valid in the continuous-time limit ($\tau = 0$).

This 'miraculous' compensation is due to a very peculiar property of the Gaussian distribution, for which:

$$P_G(x_1, T|x_0, 0)\delta(x_1 - x_2) - P_G(x_1, T|x_0, 0)P_G(x_2, T|x_0, 0) =$$
$$D\int_0^T \int P_G(x, t|x_0, 0)\frac{\partial P_G(x_1, T|x, t)}{\partial x}\frac{\partial P_G(x_2, T|x, t)}{\partial x}\,dx\,dt. \tag{4.90}$$

Integrating this equation with respect to x_1 and x_2 after multiplication by the corresponding payoff functions, the left-hand side gives $\mathcal{R}_0^2$. As for the right-hand side, one recognizes the limit of $\sum_{k=0}^{N-1} \int P(x, k|x_0, 0)[\phi_k^{N*}(x)]^2\,dx$ when $\tau = 0$, where the sum becomes an integral.

Therefore, in the continuous-time Gaussian case, the Δ-hedge of Black and Scholes allows one to eliminate completely the risk associated with an option, as was the case for the forward contract. However, contrarily to the forward contract where the elimination of risk is possible whatever the statistical nature of the fluctuations, the result of Black and Scholes is only valid in a very special limiting case. For example, as soon as the elementary time scale τ is finite (which is the case in reality), the residual risk $\mathcal{R}^*$ is non-zero even if the increments are Gaussian. The calculation is easily done in the limit where τ is small compared to T: the residual risk then comes from the difference between a continuous integral $D\tau \int dt' \int P_G(x, t'|x_0, 0)\phi^{*2}(x, t')\,dx$ (which is equal to $\mathcal{R}_0^2$) and the corresponding discrete sum appearing in Eq. (4.89). This difference is given by the Euler–McLaurin formula and is equal to:

$$\mathcal{R}^* = \sqrt{\frac{D\tau}{2}\mathcal{P}(1-\mathcal{P})} + O(\tau), \tag{4.91}$$

where $\mathcal{P}$ is the probability (at $t = 0$) that the option is exercised at expiry ($t = T$).[26] In the limit $\tau \to 0$, one indeed recovers $\mathcal{R}^* = 0$, which also holds if $\mathcal{P} \to 0$ or 1, since the outcome of the option then becomes certain. However, Eq. (4.91) already shows that in reality, the residual risk is *not small*. Take for example an at-the-money option, for which $\mathcal{P} = \frac{1}{2}$. The comparison between the residual risk and the price of the option allows one to define a 'quality' ratio Q for the hedging strategy:

$$Q \equiv \frac{\mathcal{R}^*}{\mathcal{C}} \simeq \sqrt{\frac{\pi}{4N}}, \tag{4.92}$$

with $N = T/\tau$. For an option of maturity 1 month, rehedged daily, $N \simeq 25$. Assuming Gaussian fluctuations, the quality ratio is then $Q \simeq 0.2$. In other words, the residual risk is one-fifth of the price of the option itself. Even if one rehedges every 30 min in a Gaussian world, the quality ratio is already $Q \simeq 0.05$. If the

[26] The above formula is only correct in the additive limit, but can be generalized to any Gaussian model, in particular the log-normal Black–Scholes model.

increments are not Gaussian, then $\mathcal{R}^*$ can never reach zero. This is actually rather intuitive, the presence of unpredictable price 'jumps' jeopardizes the differential strategy of Black–Scholes. The miraculous compensation of the two terms in Eq. (4.89) no longer takes place. Figure 4.11 gives the residual risk as a function of τ for an option on the Bund contract, calculated using the optimal strategy ϕ^*, and assuming independent increments. As expected, $\mathcal{R}^*$ increases with τ, but does *not* tend to zero when τ decreases: the theoretical quality ratio saturates around $Q = 0.17$. In fact, the real risk is even larger than the theoretical estimate since the model is imperfect. In particular, it neglects the volatility fluctuations, i.e. that of the scale factor γ_k. (The importance of these volatility fluctuations for determining the correct price was discussed above in Section 4.3.4.) In other words, the theoretical curve shown in Figure 4.11 neglects what is usually called the 'volatility' risk. A Monte-Carlo simulation of the profit and loss associated to an at-the-money option, hedged using the optimal strategy ϕ^* determined above leads to the histogram shown in Figure 4.8. The empirical variance of this histogram corresponds to $Q \simeq 0.28$, substantially larger than the theoretical estimate.

The 'stop-loss' strategy does not work

There is a very simple strategy that leads, at first sight, to a perfect hedge. This strategy is to hold $\phi = 1$ of the underlying as soon as the price x exceeds the strike price x_s, and to sell everything ($\phi = 0$) as soon as the price falls below x_s. For zero interest rates, this 'stop-loss' strategy would obviously lead to zero risk, since either the option is exercised at time T, but the writer of the option has bought the underlying when its value was x_s, or the option is not exercised, but the writer does not possess any stock. If this were true, the price of the option would actually be zero, since in the global wealth balance, the term related to the hedge perfectly matches the option pay-off!

In fact, this strategy does not work at all. A way to see this is to realize that when the strategy takes place in discrete time, the price of the underlying is never *exactly* at x_s, but slightly above, or slightly below. If the trading time is τ, the difference between x_s and x_k (where k is the time in discrete units) is, for a random walk, of the order of $\sqrt{D\tau}$. The difference between the ideal strategy, where the buy or sell order is always executed precisely at x_s, and the real strategy is thus of the order of $N_\times\sqrt{D\tau}$, where $N_\times$ is the number of times the price crosses the value x_s during the lifetime T of the option. For an at-the-money option, this is equal to the number of times a random walk returns to its starting point in a time T. The result is well known, and is $N_\times \propto \sqrt{T/\tau}$ for $T \gg \tau$. Therefore, the uncertainty in the final result due to the accumulation of the above small errors is found to be of order $\sqrt{DT}$, independently of τ, and

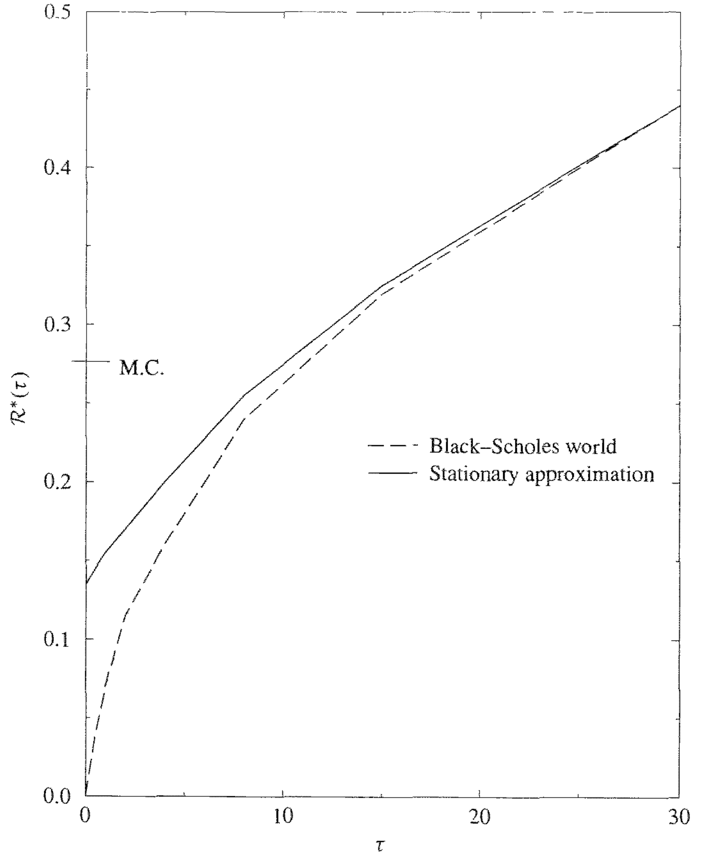

Fig. 4.11. Residual risk $\mathcal{R}^*$ as a function of the time τ (in days) between adjustments of the hedging strategy ϕ^*, for at-the-money options of maturity equal to 60 trading days. The risk $\mathcal{R}^*$ decreases when the trading frequency increases, but only rather slowly: the risk only falls by 10% when the trading time drops from 1 day to 30 min. Furthermore, $\mathcal{R}^*$ does not tend to zero when $\tau \to 0$, at variance with the Gaussian case (dotted line). Finally, the present model neglects the volatility risk, which leads to an even larger residual risk (marked by the symbol 'M.C.'), corresponding to a Monte-Carlo simulation using real price changes.

therefore does not vanish in the continuous-time limit $\tau \to 0$. This uncertainty is of the order of the option price itself, even for the case of a Gaussian process. Hence, the 'stop-loss' strategy is not the optimal strategy, and was therefore not found as a solution of the functional minimization of the risk presented above.

Residual risk to first order in kurtosis

It is interesting to compute the first non-Gaussian correction to the residual risk, which is induced by a non-zero kurtosis κ_1 of the distribution at scale τ. The expression of $\partial\mathcal{R}^*/\partial\kappa_1$ can be obtained from Eq. (4.89), where two types of term appear: those proportional to $\partial P/\partial\kappa_1$, and those proportional to $\partial\phi^*/\partial\kappa_1$. The latter terms are zero since by definition the derivative of $\mathcal{R}^*$ with respect to ϕ^* is nil. We thus find:

$$\frac{\partial \mathcal{R}^{*2}}{\partial \kappa_1} = \int_{x_s}^{\infty} (x - x_s)[(x - x_s) - 2\langle \max(x(N) - x_s, 0)\rangle] \times \frac{\partial P(x, N|x_0, 0)}{\partial \kappa_1}\,dx - D\tau \sum_k \int \frac{\partial P(x', k|x_0, 0)}{\partial \kappa_1}[\phi_k^{N*}]^2\,dx'. \tag{4.93}$$

Now, to first order in kurtosis, one has (cf. Chapter 1):

$$\frac{\partial P(x', k|x_0, 0)}{\partial \kappa_1} = k\,\frac{(D\tau)^2}{4!}\frac{\partial^4 P_G(x', k|x_0, 0)}{\partial x_0^4}, \tag{4.94}$$

which allows one to estimate the extra risk numerically, using Eq. (4.93). The order of magnitude of all the above terms is given by $\partial\mathcal{R}^{*2}/\partial\kappa_1 \simeq D\tau/4!$. The relative correction to the Gaussian result, Eq. (4.91), is then simply given, very roughly, by the kurtosis κ_1. Therefore, tail effects, as measured by the kurtosis, constitute the major source of the residual risk when the trading time τ is small.

Stochastic volatility models

It is instructive to consider the case where the fluctuations are purely Gaussian, but where the volatility itself is randomly varying. In other words, the instantaneous variance is given by $D_k = \overline{D} + \delta D_k$, where δD_k is itself a random variable of variance $\overline{(\delta D)^2}$. If the different δD_k's are uncorrelated, this model leads to a non-zero kurtosis (cf. Eq. (2.17) and Section 1.7.2) equal to $\kappa_1 = 3\overline{(\delta D)^2}/\overline{D}^2$.

Suppose then that the option trader follows the Black–Scholes strategy corresponding to the average volatility $\overline{D}$. This strategy is close, but not identical to, the optimal strategy which he would follow if he knew all the future values of δD_k (this strategy is given in Appendix D). If $\delta D_k \ll \overline{D}$, one can perform a perturbative calculation of the residual risk associated with the uncertainty on the volatility. This excess risk is certainly of order $(\delta D)^2$ since one is close to the absolute minimum of the risk, which is reached for $\delta D \equiv 0$. The calculation does indeed yield a relative increase in risk whose order of magnitude is given by:

$$\frac{\delta\mathcal{R}^*}{\mathcal{R}^*} \propto \frac{\overline{(\delta D)^2}}{\overline{D}^2}. \tag{4.95}$$

If the observed kurtosis is entirely due to this stochastic volatility effect, one has $\delta\mathcal{R}^*/\mathcal{R}^* \propto \kappa_1$. One thus recovers the result of the previous section. Again, this volatility risk can represent an appreciable fraction of the real risk, especially for frequently hedged options. Figure 4.11 actually suggests that the fraction of the risk due to the fluctuating volatility is comparable to that induced by the intrinsic kurtosis of the distribution.

4.4.6 Other measures of risk – hedging and VaR (*)

It is conceptually illuminating to consider the model where the price increments δx_k are Lévy variables of index $\mu < 2$, for which the variance is infinite. Such a model is also useful to describe extremely volatile markets, such as emergent countries (like Russia!), or very speculative assets (cf. Chapter 2). In this case, the variance of the global wealth balance ΔW is meaningless and cannot be used as a reliable measure of the risk associated to the option. Indeed, one expects that the distribution of ΔW behaves, for large negative ΔW, as a power-law of exponent μ.

In order to simplify the discussion, let us come back to the case where the hedging strategy ϕ is independent of time, and suppose interest rate effects negligible. The wealth balance, Eq. (4.28), then clearly shows that the catastrophic losses occur in two complementary cases:

- Either the price of the underlying soars rapidly, far beyond both the strike price x_s and x_0. The option is exercised, and the loss incurred by the writer is then:

$$|\Delta W_+| = x_N(1-\phi) - x_s + \phi x_0 - \mathcal{C} \simeq x_N(1-\phi). \tag{4.96}$$

  The hedging strategy ϕ, in this case, limits the loss, and the writer of the option would have been better off holding $\phi = 1$ stock per written option.
- Or, on the contrary, the price plummets far below x_0. The option is then not exercised, but the strategy leads to important losses since:

$$|\Delta W_-| = \phi(x_0 - x_N) - \mathcal{C} \simeq -\phi x_N. \tag{4.97}$$

  In this case, the writer of the option should not have held any stock at all ($\phi = 0$).

However, both dangers are *a priori* possible. Which strategy should one follow? Thanks to the above argument, it is easy to obtain the tail of the distribution of ΔW when $\Delta W \to -\infty$ (large losses). Since we have assumed that the distribution of $x_N - x_0$ decreases as a power-law for large arguments,

$$P(x_N, N|x_0, 0) \underset{x_N - x_0 \to \pm\infty}{\simeq} \frac{\mu A_\pm^\mu}{|x_N - x_0|^{1+\mu}}, \tag{4.98}$$

it is easy to show, using the results of Appendix C, that:

$$P(\Delta W) \underset{\Delta W \to -\infty}{\simeq} \frac{\mu \Delta W_0^\mu}{|\Delta W|^{1+\mu}} \qquad \Delta W_0^\mu \equiv A_+^\mu (1-\phi)^\mu + A_-^\mu \phi^\mu. \tag{4.99}$$

The probability that the loss $|\Delta W|$ is larger than a certain value is then proportional to ΔW_0^μ (cf. Chapter 3). The minimization of this probability with respect to ϕ then

leads to an optimal 'value-at-risk' strategy:

$$\phi^* = \frac{A_+^\zeta}{A_+^\zeta + A_-^\zeta} \qquad \zeta \equiv \frac{\mu}{\mu - 1}, \tag{4.100}$$

for $1 < \mu \le 2$.[27] For $\mu < 1$, ϕ^* is equal to 0 if $A_- > A_+$ or to 1 in the opposite case. Several points are worth emphasizing:

- The hedge ratio ϕ^* is independent of moneyness $(x_s - x_0)$. Because we are interested in minimizing extreme risks, only the far tail of the wealth distribution matters. We have implicitly assumed that we are interested in moves of the stock price far greater than $|x_s - x_0|$, i.e. that moneyness only influences the centre of the distribution of ΔW.
- It can be shown that within this value-at-risk perspective, the strategy ϕ^* is actually time independent, and also corresponds to the optimal instantaneous hedge, where the VaR between times k and $k+1$ is minimum.
- Even if the tail amplitude ΔW_0 is minimum, the variance of the final wealth is still infinite for $\mu < 2$. However, ΔW_0^* sets the order of magnitude of probable losses, for example with 95% confidence. As in the example of the optimal portfolio discussed in Chapter 3, infinite variance does not necessarily mean that the risk cannot be diversified. The option price, fixed for $\mu > 1$ by Eq. (4.33), ought to be corrected by a risk premium proportional to ΔW_0^*. Note also that with such violent fluctuations, the smile becomes a spike!
- Finally, it is instructive to note that the histogram of ΔW is completely asymmetrical, since extreme events only contribute to the loss side of the distribution. As for gains, they are limited, since the distribution decreases very fast for positive ΔW.[28] In this case, the analogy between an option and an insurance contract is most clear, and shows that buying or selling an option are not at all equivalent operations, as they appear to be in a Black–Scholes world. Note that the asymmetry of the histogram of ΔW is visible even in the case of weakly non-Gaussian assets (Fig. 4.8).

As we have just discussed, the losses of the writer of an option can be very large in the case of a Lévy process. Even in the case of a 'truncated' Lévy process (in the sense defined in Chapters 1 and 2), the distribution of the wealth balance ΔW remains skewed towards the loss side. It can therefore be justified to consider other measures of risk, not based on the variance but rather on higher moments of the distribution, such as $\mathcal{R}_4 = (\langle \Delta W^4 \rangle)^{1/4}$, which are more sensitive to large losses. The minimization of $\mathcal{R}_4$ with respect to $\phi(x,t)$ can still be performed, but leads

[27] Note that the above strategy is still valid for $\mu > 2$, and corresponds to the optimal VaR hedge for power-law-tailed assets, see below.
[28] For at-the-money options, one can actually show that $\Delta W \le \mathcal{C}$.

to a more complex equation for ϕ^*, which has to be solved numerically. One finds that this optimal strategy varies more slowly with the underlying price x than that based on the minimization of the variance, which is interesting from the point of view of transaction costs.[29]

Another possibility is to measure the risk through the value-at-risk (or loss probability), as is natural to do in the case of the Lévy processes discussed above. If the underlying asset has power-law fluctuations with an exponent $\mu > 2$, the above computation remains valid as long as one is concerned with the extreme tail of the loss distribution (cf. Chapter 3). The optimal VaR strategy, minimizing the probability of extreme losses, is determined by Eq. (4.100). This strategy is furthermore time independent, and therefore is very interesting from the point of view of transaction costs.

4.4.7 Hedging errors

The formulation of the hedging problem as a variational problem has a rather interesting consequence, which is a certain amount of stability against hedging errors. Suppose indeed that instead of following the optimal strategy $\phi^*(x,t)$, one uses a suboptimal strategy close to ϕ^*, such as the Black–Scholes hedging strategy with the value of the implied volatility, discussed in Section 4.4.3. Denoting the difference between the actual strategy and the optimal one by $\delta\phi(x,t)$, one can show that for small $\delta\phi$, the increase in residual risk is given by:

$$\delta\mathcal{R}^2 = D\tau \sum_{k=0}^{N-1} \int [\delta\phi(x,t_k)]^2 P(x,k|x_0,0)\,\mathrm{d}x, \tag{4.101}$$

which is *quadratic* in the hedging error $\delta\phi$, and thus, in general, rather small. For example, we have estimated in Section 4.4.3 that within a first-order cumulant expansion, $\delta\phi$ is at most equal to $0.02\kappa_N$, where κ_N is the kurtosis corresponding to the terminal distribution. (See also Fig. 4.10.) Therefore, one has:

$$\delta\mathcal{R}^2 \le 4\,10^{-4}\kappa_N^2 DT. \tag{4.102}$$

For at-the-money options, this corresponds to a relative increase of the residual risk given by:

$$\frac{\delta\mathcal{R}}{\mathcal{R}} \le 1.2\,10^{-3}\frac{\kappa_N^2}{Q^2}. \tag{4.103}$$

For a quality ratio $Q = 0.25$ and $\kappa_N = 1$, this represents a rather small relative increase equal to 2% at most. In reality, as numerical simulations show, the increase in risk induced by the use of the Black–Scholes Δ-hedge rather than the optimal hedge is indeed only of a few per cent for 1-month maturity options. This difference however increases as the maturity of the options decreases.

[29] This has been shown in the PhD work of Farhat Selmi (2000), unpublished.

4.4.8 Summary

In this part of the chapter, we have thus shown that one can find an optimal hedging strategy, in the sense that the risk (measured by the variance of the change of wealth) is minimum. This strategy can be obtained explicitly, with very few assumptions, and is given by Eqs (4.76), or (4.80). However, for a non-linear pay-off, the residual risk is in general non-zero, and actually represents an appreciable fraction of the price of the option itself. The exception is the Black–Scholes model where the risk, rather miraculously, disappears. The theory presented here generalizes that of Black and Scholes, and is formulated as a variational theory. Interestingly, this means that small errors in the hedging strategy increases the risk only in second order.

4.5 Does the price of an option depend on the mean return?

4.5.1 The case of non-zero excess return

We should now come back to the case where the excess return $m_1 \equiv \langle\delta x_k\rangle = m\tau$ is non-zero. This case is very important conceptually: indeed, one of the most striking result of Black and Scholes (besides the zero risk property) is that the price of the option and the hedging strategy are *totally independent* of the value of m. This may sound at first rather strange, since one could think that if m is very large and positive, the price of the underlying asset on average increases fast, thereby increasing the average pay-off of the option. On the contrary, if m is large and negative, the option should be worthless.

This argument actually does not take into account the impact of the hedging strategy on the global wealth balance, which is proportional to m. In other words, the term $\max(x(N) - x_s, 0)$, averaged with the historical distribution $P_m(x, N|x_0, 0)$, such that:

$$P_m(x, N|x_0, 0) = P_{m=0}(x - Nm_1, N|x_0, 0), \tag{4.104}$$

is indeed strongly dependent on the value of m. However, this dependence is partly compensated when one includes the trading strategy, and even vanishes in the Black–Scholes model.

Let us first present a perturbative calculation, assuming that m is small, or more precisely that $(mT)^2/DT \ll 1$. Typically, for $T = 100$ days, $mT = 5\%100/365 = 0.014$ and $\sqrt{DT} \simeq 1\%\sqrt{100} \simeq 0.1$. The term of order m^2 that we neglect corresponds to a relative error of $(0.14)^2 \simeq 0.02$.

The average gain (or loss) induced by the hedge is equal to:[30]

$$\langle \Delta W_S \rangle = +m_1 \sum_{k=0}^{N-1} \int P_m(x, k|x_0, 0)\phi_k^{N*}(x)\,\mathrm{d}x. \tag{4.105}$$

To order m, one can consistently use the general result Eq. (4.80) for the optimal hedge ϕ^*, established above for $m = 0$, and also replace P_m by $P_{m=0}$:

$$\begin{aligned}\langle \Delta W_S \rangle \;&=\; -m_1 \sum_{k=0}^{N-1} \int P_0(x, k|x_0, 0) \\ &\times \int_{x_s}^{+\infty} (x' - x_s) \sum_{n=2}^{\infty} \frac{(-)^n c_{n,1}}{D\tau (n-1)!} \frac{\partial^{n-1}}{\partial x'^{n-1}} P_0(x', N|x, k)\,\mathrm{d}x'\,\mathrm{d}x.\end{aligned} \tag{4.106}$$

where P_0 is the unbiased distribution ($m = 0$).

Now, using the Chapman–Kolmogorov equation for conditional probabilities:

$$\int P_0(x', N|x, k) P_0(x, k|x_0, 0)\,\mathrm{d}x = P_0(x', N|x_0, 0), \tag{4.107}$$

one easily derives, after an integration by parts, and using the fact that $P_0(x', N|x_0, 0)$ only depends on $x' - x_0$, the following expression:

$$\begin{aligned}\langle \Delta W_S \rangle \;&=\; m_1 N \left[\int_{x_s}^{+\infty} P_0(x', N|x_0, 0)\,\mathrm{d}x' \right. \\ &\left. + \sum_{n=3}^{\infty} \frac{c_{n,1}}{D\tau (n-1)!} \frac{\partial^{n-3}}{\partial x_0^{n-3}} P_0(x_s, N|x_0, 0) \right].\end{aligned} \tag{4.108}$$

On the other hand, the increase of the average pay-off, due to a non-zero mean value of the increments, is calculated as:

$$\begin{aligned}\langle \max(x(N) - x_s, 0) \rangle_m \;&\equiv\; \int_{x_s}^{\infty} (x' - x_s) P_m(x', N|x_0, 0)\,\mathrm{d}x' \\ &=\; \int_{x_s - m_1 N}^{\infty} (x' + m_1 N - x_s) P_0(x', N|x_0, 0)\,\mathrm{d}x',\end{aligned} \tag{4.109}$$

where in the second integral, the shift $x_k \to x_k + m_1 k$ allowed us to remove the bias m, and therefore to substitute P_m by P_0. To first order in m, one then finds:

$$\begin{aligned}\langle \max(x(N) - x_s, 0) \rangle_m \;&\simeq\; \langle \max(x(N) - x_s, 0) \rangle_0 + \\ &m_1 N \int_{x_s}^{\infty} P_0(x', N|x_0, 0)\,\mathrm{d}x'.\end{aligned} \tag{4.110}$$

[30] In the following, we shall again stick to an additive model and discard interest rate effects, in order to focus on the main concepts.

Hence, grouping together Eqs (4.108) and (4.110), one finally obtains the price of the option in the presence of a non-zero average return as:

$$\mathcal{C}_m = \mathcal{C}_0 - \frac{mT}{D\tau} \sum_{n=3}^{\infty} \frac{c_{n,1}}{(n-1)!} \frac{\partial^{n-3}}{\partial x_0^{n-3}} P_0(x_s, N|x_0, 0). \tag{4.111}$$

Quite remarkably, *the correction terms are zero in the Gaussian case*, since all the cumulants $c_{n,1}$ are zero for $n \geq 3$. In fact, in the Gaussian case, this property holds to all orders in m (cf. below). However, for non-Gaussian fluctuations, one finds that a non-zero return should in principle affect the price of the options. Using again Eq. (4.79), one can rewrite Eq. (4.111) in a simpler form as:

$$\mathcal{C}_m = \mathcal{C}_0 + mT\left[\mathcal{P} - \phi^*\right] \tag{4.112}$$

where $\mathcal{P}$ is the probability that the option is exercised, and ϕ^* the optimal strategy, both calculated at $t = 0$. From Fig. 4.10 one sees that in the presence of 'fat tails', a positive average return makes out-of-the-money options less expensive ($\mathcal{P} < \phi^*$), whereas in-the-money options should be more expensive ($\mathcal{P} > \phi^*$). Again, the Gaussian model (for which $\mathcal{P} = \phi^*$) is misleading:[31] the independence of the option price with respect to the market 'trend' only holds for Gaussian processes, and is no longer valid in the presence of 'jumps'. Note however that the correction is usually numerically quite small: for $x_0 = 100$, $m = 10\%$ per year, $T = 100$ days, and $|\mathcal{P} - \phi^*| \sim 0.02$, one finds that the price change is of the order of 0.05 points, while $\mathcal{C} \simeq 4$ points.

'Risk neutral' probability

It is interesting to notice that the result, Eq. (4.111), can alternatively be rewritten as:

$$\mathcal{C}_m = \int_{x_s}^{\infty} (x - x_s) Q(x, N|x_0, 0)\,\mathrm{d}x, \tag{4.113}$$

with an 'effective probability' (called 'risk neutral probability', or 'pricing kernel' in the mathematical literature) Q defined as:

$$\begin{aligned}Q(x, N|x_0, 0) \;&=\; P_0(x, N|x_0, 0) \\ &- \frac{m_1}{D\tau} \sum_{n=3}^{\infty} \frac{c_{n,N}}{(n-1)!} \frac{\partial^{n-1}}{\partial x_0^{n-1}} P_0(x', N|x_0, 0),\end{aligned} \tag{4.114}$$

which satisfies the following equations:

$$\int Q(x, N|x_0, 0)\,\mathrm{d}x = 1, \tag{4.115}$$

$$\int (x - x_0) Q(x, N|x_0, 0)\,\mathrm{d}x = 0. \tag{4.116}$$

[31] The fact that the optimal strategy is equal to the probability $\mathcal{P}$ of exercising the option also holds in the Black–Scholes model, up to small σ^2 correction terms.

Note that the second equation means that the evolution of x under the 'probability' Q is unbiased, i.e. $\langle x\rangle|_Q = x_0$. This is the definition of a martingale process (see, e.g. [Baxter]). Equation (4.113), with the very same Q, actually holds for an arbitrary pay-off function $\mathcal{Y}(x_N)$, replacing $\max(x_N - x_s, 0)$ in the above equation. Using Eq. (4.79), Eq. (4.114) can also be written in a more compact way as:

$$Q(x, N|x_0, 0) = P_0(x, N|x_0, 0) - \frac{m_1}{D\tau}(x - x_0)P_0(x, N|x_0, 0) + m_1 N \frac{\partial P_0(x, N|x_0, 0)}{\partial x_0}. \tag{4.117}$$

Note however that Eq. (4.113) is rather formal, since nothing ensures that $Q(x, N|x_0, 0)$ is everywhere positive, except in the Gaussian case where $Q(x, N|x_0, 0) = P_0(x, N|x_0, 0)$.

As we have discussed above, small errors in the hedging strategy do not significantly increase the risk. If the followed strategy is not the optimal one but, for example, the Black–Scholes 'Δ'-hedge (i.e. $\phi = \Delta = \partial C/\partial x$), the fair game price is given by Eq. (4.113) with $Q(x, N|x_0, 0) = P_0(x, N|x_0, 0)$, and is now independent of m and positive everywhere.[32] The difference between this 'suboptimal' price and the truly optimal one is then, according to Eq. (4.112), equal to $\delta C = mT(\phi^* - \mathcal{P})$. As already discussed, the above difference is quite small, and leads to price corrections which are, in most cases, negligible compared to the uncertainty in the volatility and to the residual risk $\mathcal{R}^*$.

Optimal strategy in the presence of a bias

We now give, without giving the details of the computation, the general equation satisfied by the optimal hedging strategy in the presence of a non-zero average return m, when the price fluctuations are arbitrary but uncorrelated. Assuming that $\langle\delta x_k \delta x_\ell\rangle - m_1^2 = D\tau\delta_{k,\ell}$, and introducing the unbiased variables $\chi_k \equiv x_k - m_1 k$, one gets for the optimal strategy $\phi_k^*(\chi)$ the following (involved) integral equation:

$$\begin{aligned}
&D\tau\phi_k^*(\chi) - \int_{\chi_s}^{\infty} (\chi' - \chi_s)\frac{\chi' - \chi}{N - k} P_0(\chi', N|\chi, k)\,d\chi' = \\
&-m_1 \Bigg\{ \int_{\chi_s}^{\infty} (\chi' - \chi_s)[P_0(\chi', N|\chi_0, 0) - P_0(\chi', N|\chi, k)]\,d\chi' \\
&+ \sum_{\ell=k+1}^{N-1} \int \left[\frac{\chi' - \chi}{\ell - k} + m_1\right]\phi_\ell^*(\chi')P_0(\chi', \ell|\chi, k)\,d\chi' \\
&+ \sum_{\ell=0}^{k-1} \int \left[\frac{\chi - \chi'}{k - \ell} + m_1\right]\phi_\ell^*(\chi')\frac{P_0(\chi', \ell|\chi_0, 0)P_0(\chi, k|\chi', \ell)}{P_0(\chi, k|\chi_0, 0)}\,d\chi' \\
&- m_1\overline{\phi^*}\Bigg\},
\end{aligned} \tag{4.118}$$

[32] If $\phi = \Delta$, the result $Q = P_0$ is in fact correct (as we show in Appendix F) beyond the first order in m. However, the optimal strategy is not, in general, given by the option Δ, except in the Gaussian case.

with $\chi_s = x_s - m_1 N$ and

$$\overline{\phi^*} \equiv \sum_{k=0}^{N-1} \int \phi_k^*(\chi')P_0(\chi', k|\chi_0, 0)\,d\chi'. \tag{4.119}$$

In the limit $m_1 = 0$, the right-hand side vanishes and one recovers the optimal strategy determined above. For small m, Eq. (4.118) is a convenient starting point for a perturbative expansion in m_1.

Using Eq. (4.118), one can establish a simple relation for $\overline{\phi^*}$, which fixes the correction to the 'Bachelier price' coming from the hedge: $C = \langle\max(x_N - x_s, 0)\rangle - m_1\overline{\phi^*}$.

$$\begin{aligned}
\overline{\phi^*} &= \sum_{k=0}^{N-1} \int_{\chi_s}^{\infty} (\chi' - \chi_s)\frac{\chi' - \chi}{N - k} P_0(\chi', N|\chi, k)P_0(\chi, k|\chi_0, 0)\,d\chi' \\
&\quad - m_1 \sum_{\ell=k+1}^{N-1} \int \frac{\chi' - \chi}{\ell - k}\phi_\ell^*(\chi')P_0(\chi', \ell|\chi_0, 0)\,d\chi'.
\end{aligned} \tag{4.120}$$

Replacing $\phi_\ell^*(\chi')$ by its corresponding equation $m_1 = 0$, one obtains the correct value of $\overline{\phi^*}$ to order m_1, and thus the value of C to order m_1^2 included.

4.5.2 The Gaussian case and the Black–Scholes limit

In the continuous-time Gaussian case, the solution of Eqs (4.118) and (4.119) for the optimal strategy ϕ^* happens to be completely independent of m (cf. next section on the Ito calculus). Coming back to the variable x, one finds:

$$\phi^*(x, t) = -\int_{x_s}^{\infty} \frac{1}{\sqrt{2\pi D(T - t)}}(x' - x_s)\frac{\partial}{\partial x'}\exp\left[-\frac{(x' - x)^2}{2D(T - t)}\right]dx'. \tag{4.121}$$

The average profit induced by the hedge is thus:

$$m_1\overline{\phi^*} = m\int_0^T \int \frac{1}{\sqrt{2\pi Dt}}\phi^*(x, t)\exp\left[-\frac{(x - x_0 - mt)^2}{2Dt}\right]dx\,dt. \tag{4.122}$$

Performing the Gaussian integral over x, one finds (setting $u = x' - x_s$ and $u_0 = x_0 - x_s$):

$$\begin{aligned}
m_1\overline{\phi^*} &= -m\int_0^T \int_0^{\infty} \frac{u}{\sqrt{2\pi DT}}\frac{\partial}{\partial u}\exp\left[-\frac{(u - u_0 - mt)^2}{2DT}\right]du\,dt \\
&= \int_0^T \int_0^{\infty} \frac{u}{\sqrt{2\pi DT}}\frac{\partial}{\partial t}\exp\left[-\frac{(u - u_0 - mt)^2}{2DT}\right]du\,dt,
\end{aligned}$$

or else:

$$m_1\overline{\phi^*} = \int_0^{\infty} \frac{u}{\sqrt{2\pi DT}}\left\{\exp\left[-\frac{(u - u_0 - mT)^2}{2DT}\right] - \exp\left[-\frac{(u - u_0)^2}{2DT}\right]\right\}du. \tag{4.123}$$

The price of the option in the presence of a non-zero return $m \neq 0$ is thus given, in the Gaussian case, by:

$$\begin{aligned} \mathcal{C}_m(x_0, x_s, T) &= \int_0^\infty \frac{u}{\sqrt{2\pi DT}} \exp\left[-\frac{(u-u_0-mT)^2}{2DT}\right] du - m\overline{_1\phi^*} \\ &= \int_0^\infty \frac{u}{\sqrt{2\pi DT}} \exp\left[-\frac{(u-u_0)^2}{2DT}\right] du \\ &\equiv \mathcal{C}_{m=0}(x_0, x_s, T), \end{aligned} \tag{4.124}$$

(cf. Eq. (4.43)). Hence, as announced above, the option price is indeed independent of m in the Gaussian case. This is actually a consequence of the fact that the trading strategy is fixed by $\phi^* = \partial \mathcal{C}_m/\partial x$, which is indeed correct (in the Gaussian case) even when $m \neq 0$. These results, rather painfully obtained here, are immediate within the framework of stochastic differential calculus, as originally used by Black and Scholes. It is thus interesting to pause for a moment and describe how option pricing theory is usually introduced.

Ito calculus[33]

The idea behind Ito's stochastic calculus is the following. Suppose that one has to consider a certain function $f(x, t)$, where x is a time-dependent variable. If x ~~was an~~ 'ordinary' variable, the variation Δf of the function f between time t and $t + \tau$ would be given, for small τ, by:

$$\Delta f = \frac{\partial f(x,t)}{\partial t}\tau + \frac{\partial f(x,t)}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 f(x,t)}{\partial x^2}\Delta x^2 + \cdots \tag{4.125}$$

with $\Delta x \equiv (\mathrm{d}x/\mathrm{d}t)\tau$. The order τ^2 in the above expansion looks negligible in the limit $\tau \to 0$. However, if x is a stochastic variable with independent increments, the order of magnitude of $x(t) - x(0)$ is fixed by the CLT and is thus given by $\sigma_1\sqrt{t/\tau}\,\xi$, where ξ is a Gaussian random variable of width equal to 1, and σ_1 is the RMS of Δx.

If the limit $\tau \to 0$ is to be well defined and non-trivial (i.e. such that the random variable ξ still plays a role), one should thus require that $\sigma_1 \propto \sqrt{\tau}$. Since σ_1 is the RMS of $(\mathrm{d}x/\mathrm{d}t)\tau$, this means that the order of magnitude of $\mathrm{d}x/\mathrm{d}t$ is proportional to $1/\sqrt{\tau}$. Hence the order of magnitude of $\Delta x^2 = (\mathrm{d}x/\mathrm{d}t)^2\tau^2$ is not τ^2 but τ: one should therefore keep this term in the expansion of Δf to order τ.

The crucial point of Ito's differential calculus is that if the stochastic process is a continuous-time Gaussian process, then for any small but finite time scale τ, Δx is already the result of an infinite sum of elementary increments. Therefore, one can rewrite Eq. (4.125), choosing as a new elementary time step $\tau' \ll \tau$, and sum all these τ/τ' equations to obtain Δf on the scale τ. Using the fact that for small τ, $\partial f/\partial x$ and $\partial^2 f/\partial x^2$ do not vary much, one finds:

$$\Delta x = \sum_{i=1}^{\tau/\tau'} \Delta x_i' \qquad \Delta x^2 = \sum_{i=1}^{\tau/\tau'} \Delta x_i'^2. \tag{4.126}$$

[33] The following section is obviously not intended to be rigorous.

Using again the CLT between scales $\tau' \ll \tau$ and τ, one finds that Δx is a Gaussian variable of RMS $\propto \sqrt{\tau}$. On the other hand, since Δx^2 is the sum of positive variables, it is equal to its mean $\sigma_1'^2\tau/\tau'$ plus terms of order $\sqrt{\tau/\tau'}$, where $\sigma_1'^2$ is the variance of $\Delta x'$. For consistency, $\sigma_1'^2$ must be of order τ'; we will thus set $\sigma_1'^2 \equiv D\tau'$.

Hence, in the limit $\tau' \to 0$, with τ fixed, Δx^2 in Eq. (4.125) becomes a non-random variable equal to $D\tau$, up to corrections of order $\sqrt{\tau'/\tau}$. Now, taking the limit $\tau \to 0$, one finally finds:

$$\lim_{\tau\to 0}\frac{\Delta f}{\tau} \equiv \frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f(x,t)}{\partial t} + \frac{\partial f(x,t)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{D}{2}\frac{\partial^2 f(x,t)}{\partial x^2}, \tag{4.127}$$

where $\lim_{\tau\to 0}\Delta x/\tau \equiv \mathrm{d}x/\mathrm{d}t$. Equation (4.127) means that in the double limit $\tau \to 0$, $\tau'/\tau \to 0$:

- *The second-order derivative term does not fluctuate, and thus does not depend on the specific realization of the random process. This is obviously at the heart of the possibility of finding a riskless hedge in this case.*[34]
- *Higher-order derivatives are negligible in the limit $\tau = 0$.*
- *Equation (4.127) remains valid in the presence of a non-zero bias m.*

Let us now apply the formula, Eq. (4.127) to the difference $\mathrm{d}f/\mathrm{d}t$ between $\mathrm{d}\mathcal{C}/\mathrm{d}t$ and $\phi\,\mathrm{d}x/\mathrm{d}t$. This represents the difference in the variation of the value of the portfolio of the buyer of an option, which is worth $\mathcal{C}$, and that of the writer of the option who holds $\phi(x, t)$ underlying assets. One finds:

$$\begin{aligned} \frac{\mathrm{d}f}{\mathrm{d}t} &= \phi\frac{\mathrm{d}x}{\mathrm{d}t} - \left[\frac{\partial\mathcal{C}(x,x_s,T-t)}{\partial t} + \frac{\partial\mathcal{C}(x,x_s,T-t)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}\right. \\ &\quad \left. + \frac{D}{2}\frac{\partial^2\mathcal{C}(x,x_s,T-t)}{\partial x^2}\right]. \end{aligned} \tag{4.128}$$

One thus immediately sees that if $\phi = \phi^* \equiv \partial\mathcal{C}(x, x_s, T-t)/\partial x$, the coefficient of the only random term in the above expression, namely $\mathrm{d}x/\mathrm{d}t$, vanishes. The evolution of the difference of value between the two portfolios is then known with certainty! In this case, no party would agree on the contract unless this difference remains fixed in time (we assume here, as above, that the interest rate is zero). In other words, $\mathrm{d}f/\mathrm{d}t \equiv 0$, leading to a partial differential equation for the price $\mathcal{C}$:

$$\frac{\partial\mathcal{C}(x,x_s,T-t)}{\partial t} = -\frac{D}{2}\frac{\partial^2\mathcal{C}(x,x_s,T-t)}{\partial x^2}, \tag{4.129}$$

with a 'final' boundary condition: $\mathcal{C}(x, x_s, 0) \equiv \max(x - x_s, 0)$, i.e. the value of the option at expiry, see Eq. (4.48). The solution of this equation is the above result, Eq. (4.43), obtained for $r = 0$, and the Black–Scholes strategy is obtained by taking the derivative of the price with respect to x, since this is the condition under which $\mathrm{d}x/\mathrm{d}t$ completely disappears from the game. Note also that the fact that the

[34] It is precisely for the same reason that the risk is also zero for a binomial process, where δx_k can only take two values, see Appendix E.

average return m is zero or non-zero does not appear in the above calculation. The price and the hedging strategy are therefore completely independent of the average return in this framework, a result obtained after rather cumbersome considerations above.

4.5.3 Conclusion. Is the price of an option unique?

Summarizing the above section, the Gaussian, continuous-time limit allows one to use very simple differential calculus rules, which only differ from the standard one through the appearance of a second-order *non-fluctuating* term – the so-called 'Ito correction'. The use of this calculus rule immediately leads to the two main results of Black and Scholes, namely: the existence of a riskless hedging strategy, and the fact that the value the average trend disappears from the final expressions. These two results are however not valid as soon as the hypothesis underlying Ito's stochastic calculus are violated (continuous-time, Gaussian statistics). The approach based on the global wealth balance, presented in the above sections, is by far less elegant but more general. It allows one to understand the very peculiar nature of the limit considered by Black and Scholes.

As we have already discussed, the existence of a non-zero residual risk (and more precisely of a negatively skewed distribution of the optimized wealth balance) necessarily means that the bid and ask prices of an option will be different, because the market makers will try to compensate for part of this risk. On the other hand, if the average return m is not zero, the fair price of the option explicitly depends on the optimal strategy ϕ^*, and thus of the chosen measure of risk (as was the case for portfolios, the optimal strategy corresponding to a minimal variance of the final result is different from the one corresponding to a minimum value-at-risk). The price therefore depends on the operator, of his definition of risk and of his ability to hedge this risk. In the Black–Scholes model, the price is uniquely determined since all definitions of risk are equivalent (and are all zero!). This property is often presented as a major advantage of the Gaussian model. Nevertheless, it is clear that it is precisely the existence of an ambiguity on the price that justifies the very existence of option markets![35] A market can only exist if *some* uncertainty remains. In this respect, it is interesting to note that new markets continually open, where more and more sources of uncertainty become tradable. Option markets correspond to a risk transfer: buying or selling a call are not identical operations (recall the skew in the final wealth distribution), except in the Black–Scholes world where

[35] This ambiguity is related to the residual risk, which, as discussed above, comes both from the presence of price 'jumps', and from the very uncertainty on the parameters describing the distribution of price changes ('volatility risk').

options would actually be useless, since they would be equivalent to holding a certain number of the underlying asset (given by the Δ).

4.6 Conclusion of the chapter: the pitfalls of zero-risk

The traditional approach to derivative pricing is to find an ideal hedging strategy, which perfectly duplicates the derivative contract. Its price is then, using an arbitrage argument, equal to that of the hedging strategy, and the residual risk is zero. This argument appears as such in nearly all the available books on derivatives and on the Black–Scholes model. For example, the last chapter of [Hull], called 'Review of Key Concepts', starts by the following sentence: *The pricing of derivatives involves the construction of riskless hedges from traded securities*. Although there is a rather wide consensus on this point of view, we feel that it is unsatisfactory to base a whole theory on exceptional situations: as explained above, both the continuous-time Gaussian model and the binomial model are very special models indeed. We think that it is more appropriate to start from the ingredient which allow the derivative markets to exist in the first place, namely *risk*. In this respect, it is interesting to compare the above quote from Hull to the following disclaimer, found on most Chicago Board Options Exchange documents: *Option trading involves risk*!

The idea that zero risk is the exception rather than the rule is important for a better pedagogy of financial risks in general; an adequate estimate of the residual risk – inherent to the trading of derivatives – has actually become one of the major concern of risk management (see also Sections 5.2, 5.3). The idea that the risk is zero is inadequate because zero cannot be a good approximation of anything. It furthermore provides a feeling of apparent security which can prove disastrous on some occasions. For example, the Black–Scholes strategy allows one, in principle, to hold an insurance against the fall of one's portfolio without buying a true Put option, but rather by following the associated hedging strategy. This is called an 'insurance portfolio', and was much used in the 1980s, when faith in the Black–Scholes model was at its highest. The idea is simply to sell a certain fraction of the portfolio when the market goes down. This fraction is fixed by the Black–Scholes Δ of the virtual option, where the strike price is the security level below which the investor does not want to plummet. During the 1987 crash, this strategy has been particularly inefficient: not only because crash situations are the most extremely non-Gaussian events that one can imagine (and thus the zero-risk idea is totally absurd), but also because this strategy feeds back onto the market to make it crash further (a drop of the market mechanically triggers further sell orders). According to the Brady commission, this mechanism has indeed significantly contributed to enhance the amplitude of the crash (see the discussion in [Hull]).

4.7 Appendix D: computation of the conditional mean

On many occasions in this chapter, we have needed to compute the mean value of the instantaneous increment δx_k, restricted on trajectories starting from x_k and ending at x_N. We assume that the δx_k's are identically distributed, up to a scale factor γ_k. In other words:

$$P_{1k}(\delta x_k) \equiv \frac{1}{\gamma_k} P_{10}\left(\frac{\delta x_k}{\gamma_k}\right). \tag{4.130}$$

The quantity we wish to compute is then:

$$P(x_N, N|x_k, k)\langle \delta x_k \rangle_{(x_k,k)\to(x_N,N)} = \tag{4.131}$$
$$\int \delta x_k\, \delta\left(x_N - x_k - \sum_{j=k}^{N-1} \delta x_j\right)\left[\prod_{j=k}^{N-1} P_{10}\left(\frac{\delta x_j}{\gamma_j}\right)\frac{\mathrm{d}\delta x_j}{\gamma_j}\right],$$

where the δ function insures that the sum of increments is indeed equal to $x_N - x_k$. Using the Fourier representation of the δ function, the right-hand side of this equation reads:

$$\frac{1}{2\pi}\int \mathrm{e}^{\mathrm{i}z(x_N-x_k)} \mathrm{i}\gamma_k \hat{P}'_{10}(\gamma_k z)\left[\prod_{j=k+1}^{N-1} \hat{P}_{10}(z\gamma_j)\right]\mathrm{d}z. \tag{4.132}$$

- In the case where all the γ_k's are equal to γ_0, one recognizes:

$$\frac{1}{2\pi}\int \mathrm{e}^{\mathrm{i}z(x_N-x_k)} \frac{\mathrm{i}}{N-k}\frac{\partial}{\partial z}[\hat{P}_{10}(z\gamma_0)]^{N-k}\,\mathrm{d}z. \tag{4.133}$$

Integrating by parts and using the fact that:

$$P(x_N, N|x_k, k) \equiv \frac{1}{2\pi}\int \mathrm{e}^{\mathrm{i}z(x_N-x_k)}[\hat{P}_{10}(z\gamma_0)]^{N-k}\,\mathrm{d}z, \tag{4.134}$$

one finally obtains the expected result:

$$\langle \delta x_k \rangle_{(x_k,k)\to(x_N,N)} \equiv \frac{x_N - x_k}{N-k}. \tag{4.135}$$

- In the case where the γ_k's are different from one another, one can write the result as a cumulant expansion, using the cumulants $c_{n,1}$ of the distribution P_{10}. After a simple computation, one finds:

$$P(x_N, N|x_k, k)\langle \delta x_k \rangle_{(x_k,k)\to(x_N,N)} =$$
$$\sum_{n=2}^{\infty} \frac{(-\gamma_k)^n c_{n,1}}{(n-1)!}\frac{\partial^{n-1}}{\partial x_N^{n-1}} P(x_N, N|x_k, k), \tag{4.136}$$

which allows one to generalize the optimal strategy in the case where the volatility is time dependent. In the Gaussian case, all the c_n for $n \geq 3$ are zero, and the last expression boils down to:

$$\langle \delta x_k \rangle_{(x_k,k)\to(x_N,N)} = \frac{\gamma_k^2(x_N - x_k)}{\sum_{j=k}^{N-1}\gamma_j^2}. \tag{4.137}$$

This expression can be used to show that the optimal strategy is indeed the Black–Scholes Δ, for arbitrary Gaussian increments in the presence of a non-zero interest rate. Suppose that

$$x_{k+1} - x_k = \rho x_k + \delta x_k, \tag{4.138}$$

where the δx_k are identically distributed. One can then write x_N as:

$$x_N = x_0(1+\rho)^N + \sum_{k=0}^{N-1} \delta x_k (1+\rho)^{N-k-1}. \tag{4.139}$$

The above formula, Eq. (4.137), then reads:

$$\langle \delta x_k \rangle_{(x_k,k)\to(x_N,N)} = \frac{\gamma_k^2(x_N - x_k(1+\rho)^{N-k})}{\sum_{j=k}^{N-1}\gamma_j^2}. \tag{4.140}$$

with $\gamma_k = (1+\rho)^{N-k-1}$.

4.8 Appendix E: binomial model

The binomial model for price evolution is due to Cox, Ross and Rubinstein, and shares with the continuous-time Gaussian model the zero risk property. This model is very much used in practice [Hull], due to its easy numerical implementation. Furthermore, the zero risk property appears in the clearest fashion. Suppose indeed that between $t_k = k\tau$ and t_{k+1}, the price difference can only take two values: $\delta x_k = \delta x_{1,2}$. For this very reason, the option value can only evolve along two paths only, to be worth (at time t_{k+1}) $\mathcal{C}_{1,2}^{k+1}$. Consider now the hedging strategy where one holds a fraction ϕ of the underlying asset, and a quantity B of bonds, with a risk-free interest rate ρ. If one chooses ϕ and B such that:

$$\phi_k(x_k + \delta x_1) + B_k(1+\rho) = \mathcal{C}_1^{k+1}, \tag{4.141}$$

$$\phi_k(x_k + \delta x_2) + B_k(1+\rho) = \mathcal{C}_2^{k+1}. \tag{4.142}$$

or else:

$$\phi_k = \frac{\mathcal{C}_1^{k+1} - \mathcal{C}_2^{k+1}}{\delta x_1 - \delta x_2}, \qquad B_k(1+\rho) = \frac{\delta x_1 \mathcal{C}_2^{k+1} - \delta x_2 \mathcal{C}_1^{k+1}}{\delta x_1 - \delta x_2}, \tag{4.143}$$

one sees that in the two possible cases ($\delta x_k = \delta x_{1,2}$), the value of the hedging portfolio is *strictly equal* to the option value. The option value at time t_k is thus equal to $C^k(x_k) = \phi_k x_k + B_k$, independently of the probability p [resp. $1-p$] that $\delta x_k = \delta x_1$ [resp. δx_2]. One then determines the option value by iterating this procedure from time $k+1 = N$, where C^N is known, and equal to $\max(x_N - x_s, 0)$. It is however easy to see that as soon as δx_k can take three or more values, it is impossible to find a perfect strategy.[36]

The Ito process can be obtained as the continuum limit of the binomial tree. But even in its discrete form, the binomial model shares some important properties with the Black–Scholes model. The independence of the premium on the transition probabilities is the analogue of the independence of the premium on the excess return m in the Black–Scholes model. The magic of zero risk in the binomial model can therefore be understood as follows. Consider the quantity $s^2 = (\delta x_k - \langle \delta x_k \rangle)^2$; s^2 is in principle random, but since changing the probabilities does not modify the option price one can pick $p = \frac{1}{2}$, making s^2 non-fluctuating ($s^2 = (\delta x_1 - \delta x_2)^2/4$). The Ito process shares this property: in the continuum limit quadratic quantities do not fluctuate. For example, the quantity

$$S^2 = \lim_{\tau \to 0} \sum_{k=0}^{T/\tau - 1} (x((k+1)\tau) - x(k\tau) - m\tau)^2, \tag{4.144}$$

is equal to DT with probability one when $x(t)$ follows an Ito process. In a sense, continuous-time Brownian motion represents a very weak form of randomness since quantities such as S^2 can be known with certainty. But it is precisely this property that allows for zero risk in the Black–Scholes world.

4.9 Appendix F: option price for (suboptimal) Δ-hedging

If $\phi = \Delta$, the 'risk neutral' probability $Q(x, N|x_0, 0)$ is simply equal to $P_0(x, N|x_0, 0)$, for N large, beyond the first order in m, as we show now. Taking $\phi = \partial C/\partial x$ leads to an implicit equation for C:

$$\begin{aligned} C(x_0, x_s, N) &= \int \mathcal{Y}(x' - x_s) P_m(x', N|x_0, 0)\, dx' \\ &\quad - m_1 \sum_{k=0}^{N-1} \int \frac{\partial C(x, x_s, N-k)}{\partial x} P_m(x, k|x_0, 0)\, dx, \end{aligned} \tag{4.145}$$

36 For a recent analysis along the lines of the present book, see E. Aurell, S. Simdyankin, Pricing Risky Options Simply, *International Journal of Theoretical and Applied Finance*, 1, 1 (1998). See also M. Schweizer, Risky options simplified, *International Journal of Theoretical and Applied Finance*, 2, 59 (1999).

with $\mathcal{Y}$ representing the pay-off of the option. This equation can be solved by making the following ansatz for C:

$$C(x, x_s, N-k) = \int \Omega(x' - x_s, N-k) P_m(x', N|x, k)\, dx' \tag{4.146}$$

where Ω is an unknown kernel which we try to determine. Using the fact that the option price only depends on the price difference between the strike price and the present price of the underlying, this gives:

$$\begin{aligned} &\int \Omega(x' - x_s, N) P_m(x', N|x_0, N)\, dx' = \\ &\int \mathcal{Y}(x' - x_s) P_m(x', N|x_0, 0)\, dx' + m_1 \sum_{k=0}^{N-1} \int P_m(x, k|x_0, 0) \\ &\times \frac{\partial}{\partial x_s} \int \Omega(x' - x_s, N-k) P_m(x', N|x, k)\, dx'\, dx. \end{aligned} \tag{4.147}$$

Now, using the Chapman–Kolmogorov equation:

$$\int P_m(x', N|x, k) P_m(x, k|x_0, 0)\, dx = P_m(x', N|x_0, 0), \tag{4.148}$$

one obtains the following equation for Ω (after changing $k \to N-k$):

$$\Omega(x' - x_s, N) + m_1 \sum_{k=1}^{N} \frac{\partial \Omega(x' - x_s, k)}{\partial x'} = \mathcal{Y}(x' - x_s). \tag{4.149}$$

The solution to this equation is $\Omega(x' - x_s, k) = \mathcal{Y}(x' - x_s - m_1 k)$. Indeed, if this is the case, one has:

$$\frac{\partial \Omega(x' - x_s, k)}{\partial x'} \equiv -\frac{1}{m_1} \frac{\partial \Omega(x' - x_s, k)}{\partial k} \tag{4.150}$$

and therefore:

$$\mathcal{Y}(x' - x_s - m_1 N) - \sum_{k=1}^{N} \frac{\partial \mathcal{Y}(x' - x_s - m_1 k)}{\partial k} \simeq \mathcal{Y}(x' - x_s) \tag{4.151}$$

where the last equality holds in the small τ, large N limit, when the sum over k can be approximated by an integral.[37] Therefore, the price of the option is given by:

$$C = \int \mathcal{Y}(x' - x_s - m_1 N) P_m(x', N|x_0, 0)\, dx'. \tag{4.152}$$

37 The resulting error is of the order of $m^2\tau/D$.

Now, using the fact that $P_m(x', N|x_0, 0) \equiv P_0(x' - m_1 N, N|x_0, 0)$, and changing variable from $x' \to x' - m_1 N$, one finally finds:

$$\mathcal{C} = \int \mathcal{Y}(x' - x_s) P_0(x', N|x_0, 0)\,\mathrm{d}x', \tag{4.153}$$

thereby proving that the pricing kernel Q is equal to P_0 if the chosen hedge is the Δ. Note that, interestingly, this is true independently of the particular pay-off function $\mathcal{Y}$: Q thus has a rather general status.

4.10 References

Some classics

L. Bachelier, *Théorie de la spéculation*, Gabay, Paris, 1995.
F. Black, M. Scholes, The pricing of options and corporate liabilities, *Journal of Political Economy*, **3**, 637 (1973).
R. C. Merton, Theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4**, 141 (1973).
J. C. Cox, S. A. Ross, M. Rubinstein, Option pricing: a simplified approach, *Journal of Financial Economics*, **7**, 229 (1979).

Market efficiency

P. H. Cootner, *The Random Character of Stock Market Prices*, MIT Press, Cambridge, 1964.
E. F. Fama, Efficient capital markets: a review of theory and empirical work, *Journal of Finance*, **25**, 383 (1970).

Optimal filters

N. Wiener, *Extrapolation, Interpolation and Smoothing of Time Series*, Wiley, New York, 1949.
A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd edn, McGraw-Hill, New York, 1991.

Options and futures

J. C. Hull, *Futures, Options and Other Derivatives*, Prentice Hall, Upper Saddle River, NJ, 1997.
P. Wilmott, J. N. Dewynne, S. D. Howison, *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press, Oxford, 1993.

Stochastic differential calculus and derivative pricing

M. Baxter, A. Rennie, *Financial Calculus, An Introduction to Derivative Pricing*, Cambridge University Press, Cambridge, 1996.
M. Musiela, M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer, New York, 1997.
N. Taleb, *Dynamical Hedging: Managing Vanilla and Exotic Options*, Wiley, New York, 1998.

Option pricing in the presence of residual risk

H. Follmer, D. Sondermann, Hedging of non redundant contigent claims, in *Essays in Honor of G. Debreu*, (W. Hildenbrand, A. Mas-Collel Etaus, eds), North-Holland, Amsterdam, 1986.
M. Schäl, On quadratic cost criteria for option hedging, *The Mathematics of Operations Research*, **19**, 121–131 (1994).
J.-P. Bouchaud, D. Sornette, The Black–Scholes option pricing problem in mathematical finance: generalization and extensions for a large class of stochastic processes, *Journal de Physique I*, **4**, 863–881 (1994).
M. Schweizer, Variance–optimal hedging in discrete time, *The Mathematics of Operations Research*, **20**, 1–32 (1995).

Kurtosis and implied cumulants

R. Jarrow, A. Rudd, Approximate option valuation for arbitrary stochastic processes, *Journal of Financial Economics*, **10**, 347 (1982).
C. J. Corrado, T. Su, Skewness and kurtosis in S&P500 index returns implied by option prices, *Journal of Financial Research*, **XIX**, 175 (1996).
J. C. Jackwerth, M. Rubinstein, Recovering probability distributions from option prices, *Journal of Finance*, **51**, 1611.
M. Potters, R. Cont, J.-P. Bouchaud, Financial markets as adaptative ecosystems, *Europhysics Letters*, **41**, 239 (1998).

Stochastic volatility models

J. C. Hull, A. White, The pricing of options on assets with stochastic volatility, *Journal of Finance*, **XLII**, 281 (1987); An analysis of the bias in option pricing caused by stochastic volatility, *Advances in Futures and Option Research*, **3**, 29 (1988).

5

Options: some more specific problems

This chapter can be skipped at first reading.

(J.-P. Bouchaud, M. Potters, *Theory of Financial Risks*.)

5.1 Other elements of the balance sheet

We have until now considered the simplest possible option problem, and tried to extract the fundamental ideas associated to its pricing and hedging. In reality, several complications appear, either in the very definition of the option contract (see next section), or in the elements that must be included in the wealth balance – for example dividends or transaction costs – that we have neglected up to now.

5.1.1 Interest rate and continuous dividends

The influence of interest rates (and continuous dividends) can be estimated using different models for the statistics of price increments. These models give slightly different answers; however, *in a first approximation*, they provide the following answer for the option price:

$$\mathcal{C}(x_0, x_s, T, r) = e^{-rT}\mathcal{C}(x_0 e^{rT}, x_s, T, r = 0), \tag{5.1}$$

where $\mathcal{C}(x_0, x_s, T, r = 0)$ is simply given by the average of the pay-off over the terminal price distribution. In the presence of a non-zero continuous dividend d, the quantity $x_0 e^{rT}$ should be replaced by $x_0 e^{(r-d)T}$, i.e. in all cases the present price of the underlying should be replaced by the corresponding *forward* price – see Eq. (4.23).

Let us present three different models leading to the above approximate formula; but with slightly different corrections to it in general. These three models offer alternative ways to understand the influence of a non-zero interest rate (and/or dividend) on the price of the option.

Systematic drift of the price

In this case, one assumes that the price evolves according to $x_{k+1} - x_k = (\rho - \delta)x_k + \delta x_k$, where δ is the dividend over a unit time period τ: $\delta = d\tau$ and δx_k are independent random variables. Note that this 'dividend' is, in the case of currency options, the interest rate corresponding to the underlying currency (and ρ is the interest rate of the reference currency).

This model is convenient because if $\langle \delta x_k \rangle$ is zero, the average cost of the hedging strategy is zero, since it includes the terms $x_{k+1} - x_k - (\rho - \delta)x_k$. However, the terminal distribution of x_N must be constructed by noting that:

$$x_N = x_0(1 + \rho - \delta)^N + \sum_{k=0}^{N-1} \delta x_k (1 + \rho - \delta)^{N-k-1}. \tag{5.2}$$

Thus the terminal distribution is, for large N, a function of the difference $x(T) - x_0 e^{(r-d)T}$. Furthermore, even if the δx_k are independent random variables of variance $D\tau$, the variance $c_2(T)$ of the terminal distribution is given by:

$$c_2(T) = \frac{D}{2(r-d)}[e^{2(r-d)T} - 1], \tag{5.3}$$

which is equal to $c_2(T) = DT(1 + (r - d)T)$ when $(r - d)T \to 0$. There is thus an increase of the effective volatility to be used in the option pricing formula within this model. This increase depends on the maturity; its relative value is, for $T = 1$ year, $r - d = 5\%$, equal to $(r - d)T/2 = 2.5\%$. (Note indeed that $c_2(T)$ is the *square* of the volatility). Up to this small change of volatility, the above rule, Eq. (5.1) is therefore appropriate.

Independence between price increments and interest rates–dividends

The above model is perhaps not very realistic since it assumes a direct relation between price changes and interest rates–dividends. It might be more appropriate to consider the case where $x_{k+1} - x_k = \delta x_k$, as the other extreme case; reality is presumably in between these two limiting models. Now the terminal distribution is a function of the difference $x_N - x_0$, with no correction to the variance brought about by interest rates. However, in the present case, the cost of the hedging strategy cannot be neglected and reads:

$$\langle \Delta W_S \rangle = -(\rho - \delta) \sum_{k=0}^{N-1} \langle x_k \phi_k^{N*}(x_k) \rangle. \tag{5.4}$$

Writing $x_k = x_0 + \sum_{j=0}^{k-1} \delta x_j$, we find that $\langle \Delta W_S \rangle$ is the sum of a term proportional to x_0 and the remainder. The first contribution therefore reads:

$$\langle \Delta W_S \rangle_1 = -(\rho - \delta) x_0 \sum_{k=0}^{N-1} \int P(x, k|x_0, 0) \phi_k^{N*}(x)\, dx. \tag{5.5}$$

This term thus has exactly the same shape as in the case of a non-zero average return treated above, Eq. (4.105), with an effective mean return $m_{1\text{eff}} = -(\rho - \delta)x_0$. If we choose for ϕ^* the suboptimal Δ-hedge (which we know does only induce minor errors), then, as shown above, this hedging cost can be reabsorbed in a change of the terminal distribution, where $x_N - x_0$ becomes $x_N - x_0 - m_{1\text{eff}}N$, or else, to first order: $x_N - x_0 \exp[(\rho - \delta)N]$. To order $(\rho - \delta)N$, this therefore again corresponds to replacing the current price of the underlying by its forward price. For the second type of term, we write, for $j < k$:

$$\langle \delta x_j \phi_k^{N*}(x_k) \rangle = \int P(x_j, j|x_0, 0)\, dx_j \int P(x_k, k|x_j, j) \frac{x_k - x_j}{k - j} \phi_k^{N*}(x_k)\, dx_k. \tag{5.6}$$

This type of term is not easy to calculate in the general case. Assuming that the process is Gaussian allows us to use the identity:

$$P(x_k, k|x_j, j) \frac{x_k - x_j}{k - j} = -D\tau \frac{\partial P(x_k, k|x_j, j)}{\partial x_k}; \tag{5.7}$$

one can therefore integrate over x_j to find a result independent of j:

$$\langle \delta x_j \phi_k^{N*}(x_k) \rangle = D\tau \frac{\partial}{\partial x_0} \int P(x_k, k|x_0, 0) \phi_k^{N*}(x_k)\, dx_k, \tag{5.8}$$

or, using the expression of $\phi_k^{N*}(x_k)$ in the Gaussian case:

$$\langle \delta x_j \phi_k^{N*}(x_k) \rangle = D\tau \frac{\partial}{\partial x_0} \int_{x_s}^{\infty} P(x', N|x_0, 0)\, dx' = D\tau P(x_s, N|x_0, 0). \tag{5.9}$$

The contribution to the cost of the strategy is then obtained by summing over k and over $j < k$, finally leading to:

$$\langle \Delta W_S \rangle_2 \simeq -\frac{N^2}{2} (\rho - \delta) D\tau P(x_s, N|x_0, 0). \tag{5.10}$$

This extra cost is maximum for at-the-money options, and leads to a relative increase of the call price given, for Gaussian statistics, by:

$$\frac{\delta C}{C} = \frac{N}{2}(\rho - \delta) = \frac{T}{2}(r - d), \tag{5.11}$$

which is thus quite small for short maturities. (For $r = 5\%$ annual, $d = 0$, and $T = 100$ days, one finds $\delta C/C \le 1\%$.)

Multiplicative model

Let us finally assume that $x_{k+1} - x_k = (\rho - \delta + \eta_k) x_k$, where the η_k's are independent random variables of zero mean. Therefore, the average cost of the strategy is zero. Furthermore, if the elementary time step τ is small, the terminal value x_N can be written as:

$$\log\left(\frac{x_N}{x_0}\right) = \sum_{k=0}^{N-1} \log(1 + \rho - \delta + \eta_k) \simeq N(\rho - \delta) + y; \qquad y = \sum_{k=0}^{N-1} \eta_k, \tag{5.12}$$

thus $x_N = x_0 e^{(r-d)T+y}$. Introducing the distribution $P(y)$, the price of the option can be written as:

$$\mathcal{C}(x_0, x_s, T, r, d) = e^{-rT} \int P(y) \max(x_0 e^{(r-d)T+y} - x_s, 0)\, dy, \tag{5.13}$$

which is obviously of the general form, Eq. (5.1).

We thus conclude that the simple rule, Eq. (5.1), above is, for many purposes, sufficient to account for interest rates and (continuous) dividends effects. A more accurate formula, including corrections of order $(r - d)T$ depends somewhat on the chosen model.

5.1.2 Interest rates corrections to the hedging strategy

It is also important to estimate the interest rate corrections to the optimal hedging strategy. Here again, one could consider the above models, that is, the systematic drift model, the independent model or the multiplicative model. In the former case, the result for a general terminal price distribution is rather involved, mainly due to the appearance of the conditional average detailed in Appendix D, Eq. (4.140). In the case where this distribution is Gaussian, the result is simply Black–Scholes' Δ hedge, i.e. the optimal strategy is the derivative of the option price with respect to the price of the underlying contract (see Eq. (4.78)). As discussed above, this simple rule leads in general to a suboptimal strategy, but the relative increase of risk is rather small. The Δ-hedge procedure, on the other hand, has the advantage that the price is independent of the average return (see Appendix F).

In the 'independent' model, where the price change is unrelated to the interest rate and/or dividend, the order ρ correction to the optimal strategy is easier to estimate in general. From the general formula, Eq. (4.118), one finds:

$$\begin{aligned}
\phi_k^*(x) &= (1+\rho)^{1+k-N} \phi_k^{*0}(x) - \frac{r-d}{D} x \mathcal{C}(x, x_s, N-k) \\
&+ \frac{r-d}{2D} x \sum_{\ell<k} \int \frac{x - x'}{k - \ell} \frac{P(x, k|x', \ell) P(x', \ell|x_0, 0)}{P(x, k|x_0, 0)} \phi_\ell^{*0}(x')\, dx' \\
&+ \frac{r-d}{2D} \sum_{\ell>k} \int \frac{x' - x}{\ell - k} x' P(x', \ell|x, k) \phi_\ell^{*0}(x')\, dx'
\end{aligned} \tag{5.14}$$

where ϕ_k^{*0} is the optimal strategy for $r = d = 0$.

Finally, for the multiplicative model, the optimal strategy is given by a much simpler expression:

$$\phi_k^*(x) = \frac{(1+\rho)^{1+k-N}}{x\sigma^2(N-k)} \times \int_{x_s}^{\infty} (x'-x_s)\log\left(\frac{x'}{x(1+\rho-\delta)^{N-k}}\right) P(x',N|x,k)\,\mathrm{d}x'. \tag{5.15}$$

5.1.3 Discrete dividends

More generally, for an arbitrary dividend d_k (per share) at time t_k, the extra term in the wealth balance reads:

$$\Delta W_D = \sum_{k=1}^{N} \phi_k^N(x_k) d_k. \tag{5.16}$$

Very often, this dividend occurs once a year, at a given date k_0: $d_k = d_0\delta_{k,k_0}$. In this case, the corresponding share price decreases immediately by the same amount (since the value of the company is decreased by an amount equal to d_0 times the number of outstanding shares): $x \to x - d_0$. Therefore, the net balance $d_k + \delta x_k$ associated to the dividend is zero. For the same reason, the probability $P_{d_0}(x, N|x_0, 0)$ is given, for $N > k_0$, by $P_{d_0=0}(x + d_0, N|x_0, 0)$. The option price is then equal to: $\mathcal{C}_{d_0}(x, x_s, N) \equiv \mathcal{C}(x, x_s + d_0, N)$. If the dividend d_0 is not known in advance with certainty, this last equation should be averaged over a distribution $P(d_0)$ describing as well as possible the probable values of d_0. A possibility is to choose the distribution of least information (maximum entropy) such that the average dividend is fixed to $\overline{d}$, which in this case reads:

$$P(d_0) = \frac{1}{\overline{d}} \exp(-d_0/\overline{d}). \tag{5.17}$$

5.1.4 Transaction costs

The problem of transaction costs is important: the rebalancing of the optimal hedge as time passes induces extra costs that must be included in the wealth balance as well. These 'costs' are of different nature – some are proportional to the number of traded shares, whereas another part is fixed, independent of the total amount of the operation. Let us examine the first situation, assuming that the rebalancing of the hedge takes place at every time step τ. The corresponding cost in then equal to:

$$\delta W_{\mathrm{tr}k} = \nu|\phi_{k+1}^* - \phi_k^*|, \tag{5.18}$$

where ν is a measure of the transaction costs. In order to keep the discussion simple, we shall assume that:

- the most important part in the variation of ϕ_k^* is due to the change of price of the underlying, and not from the explicit time dependence of ϕ_k^* (this is justified in the limit where $\tau \ll T$);
- δx_k is small enough such that a Taylor expansion of ϕ^* is acceptable;
- the kurtosis effects on ϕ^* are neglected, which means that the simple Black–Scholes Δ hedge is acceptable and does not lead to large hedging errors.

One can therefore write that:

$$\delta W_{\mathrm{tr}k} \doteq \nu \left|\delta x_k \frac{\partial \phi_k^*(x)}{\partial x}\right| = \nu|\delta x_k| \frac{\partial \phi_k^*(x)}{\partial x}, \tag{5.19}$$

since $\partial\phi_k^*(x)/\partial x$ is positive. The average total cost associated to rehedging is then, after a straightforward calculation:[1]

$$\langle \Delta W_{\mathrm{tr}} \rangle = \left\langle \sum_{k=0}^{N-1} \delta W_{\mathrm{tr}k} \right\rangle = \nu \langle|\delta x|\rangle N P(x_s, N|x_0, 0). \tag{5.20}$$

The order of magnitude of $\langle|\delta x|\rangle$ is given by $\sigma_1 x_0$: for an at-the-money option, $P(x_s, N|x_0, 0) \simeq (\sigma_1 x_0 \sqrt{N})^{-1}$; hence, finally, $\langle \Delta W_{\mathrm{tr}} \rangle \simeq \nu\sqrt{N}$. It is natural to compare $\langle \Delta W_{\mathrm{tr}} \rangle$ to the option price itself, which is of order $\mathcal{C} \simeq \sigma_1 x_0 \sqrt{N}$:

$$\frac{\langle \Delta W_{\mathrm{tr}} \rangle}{\mathcal{C}} \propto \frac{\nu}{\sigma_1 x_0}. \tag{5.21}$$

This part of the transaction costs is in general proportional to x_0: taking for example $\nu = 10^{-4}x_0$, $\tau = 1$ day, and a daily volatility of $\sigma_1 = 1\%$, one finds that the transaction costs represent 1% of the option price. On the other hand, for higher transaction costs (say $\nu = 10^{-2}x_0$), a daily rehedging becomes absurd, since the ratio, Eq. (5.21), is of order 1. The rehedging frequency should then be lowered, such that the volatility on the scale of τ increases, to become smaller than ν.

The fixed part of the transaction costs is easier to discuss. If these costs are equal to ν' per transaction, and if the hedging strategy is rebalanced every τ, the total cost incurred is simply given by:

$$\Delta W'_{\mathrm{tr}} = N\nu'. \tag{5.22}$$

Comparing the two types of costs leads to:

$$\frac{\Delta W'_{\mathrm{tr}}}{\langle \Delta W_{\mathrm{tr}} \rangle} \propto \frac{\nu'\sqrt{N}}{\nu}, \tag{5.23}$$

showing that the latter costs can exceed the former when $N = T/\tau$ is large, i.e. when the hedging frequency is high.

[1] One should add to the following formula the cost associated with the initial value of the hedge ϕ_0^*.

In summary, the transaction costs are higher when the trading time is smaller. On the other hand, decreasing of τ allows one to diminish the residual risk (Fig. 4.11). This analysis suggests that the optimal trading frequency should be chosen such that the transaction costs are comparable to the residual risk.

5.2 Other types of options: 'Puts' and 'exotic options'

5.2.1 'Put–call' parity

A 'Put' contract is a sell option, defined by a pay-off at maturity given by $\max(x_s - x, 0)$. A put protects its owner against the risk that the shares he owns drops below the strike price x_s. Puts on stock indices like the S&P 500 are very popular. The price of a European put will be noted $\mathcal{C}^{\dagger}[x_0, x_s, T]$ (we reserve the notation $\mathcal{P}$ for a probability). This price can be obtained using a very simple 'parity' (or no arbitrage) argument. Suppose that one simultaneously buys a call with strike price x_s, and a put with the same maturity and strike price. At expiry, this combination is therefore worth:

$$\max(x(T) - x_s, 0) - \max(x_s - x(T), 0) \equiv x(T) - x_s. \tag{5.24}$$

Exactly the same pay-off can be achieved by buying the underlying now, and selling a bond paying x_s at maturity. Therefore, the above call+put combination must be worth:

$$\mathcal{C}[x_0, x_s, T] - \mathcal{C}^{\dagger}[x_0, x_s, T] = x_0 - x_s e^{-rT}, \tag{5.25}$$

which allows one to express the price of a put knowing that of a call. If interest rate effects are small ($rT \ll 1$), this relation reads:

$$\mathcal{C}^{\dagger}[x_0, x_s, T] \simeq \mathcal{C}[x_0, x_s, T] + x_s - x_0. \tag{5.26}$$

Note in particular that at-the-money ($x_s = x_0$), the two contracts have the same value, which is obvious by symmetry (again, in the absence of interest rate effects).

5.2.2 'Digital' options

More general option contracts can stipulate that the pay-off is not the difference between the value of the underlying at maturity x_N and the strike price x_s, but rather an arbitrary function $\mathcal{Y}(x_N)$ of x_N. For example, a 'digital' option is a pure bet, in the sense that it pays a fixed premium $\mathcal{Y}_0$ whenever x_N exceeds x_s. Therefore:

$$\mathcal{Y}(x_N) = \mathcal{Y}_0 \quad (x_N > x_s); \qquad \mathcal{Y}(x_N) = 0 \quad (x_N < x_s). \tag{5.27}$$

The price of the option in this case can be obtained following the same lines as above. In particular, in the absence of bias (i.e. for $m = 0$) the fair price is given

by:

$$\mathcal{C}_{\mathcal{Y}}(x_0, N) = \langle \mathcal{Y}(x_N) \rangle = \int \mathcal{Y}(x) P_0(x, N|x_0, 0)\, dx,$$

whereas the optimal strategy is still given by the general formula, Eq. (4
particular, for Gaussian fluctuations, one always finds the Black–Scholes r

$$\phi^*_{\mathcal{Y}}(x) \equiv \frac{\partial \mathcal{C}_{\mathcal{Y}}[x, N]}{\partial x}.$$

The case of a non-zero average return can be treated as above, in Section 4.5.1 order in m, the price of the option is given by:

$$\mathcal{C}_{\mathcal{Y},m} = \mathcal{C}_{\mathcal{Y},m=0} - \frac{mT}{D\tau} \sum_{n=3}^{\infty} \frac{c_{n,1}}{(n-1)!} \int \mathcal{Y}(x') \frac{\partial^{n-1}}{\partial x^{n-1}} P_0(x', N|x, 0)\, dx',$$

which reveals, again, that in the Gaussian case, the average return disappears final formula. In the presence of non-zero kurtosis, however, a (small) systematic c to the fair price appears. Note that $\mathcal{C}_{\mathcal{Y},m}$ can be written as an average of $\mathcal{Y}(x)$ effective, 'risk neutral' probability $Q(x)$ introduced in Section 4.5.1. If the Δ- used, this risk neutral probability is simply P_0 (see Appendix F).

5.2.3 'Asian' options

The problem is slightly more complicated in the case of the so-called options. The pay-off of these options is calculated not on the value underlying stock at maturity, but on a certain average of this value over a number of days preceeding maturity. This procedure is used to prevent an a rise of the stock price precisely on the expiry date, a rise that could be tr by an operator having an important long position on the corresponding opti contract is thus constructed on a fictitious asset, the price of which being as:

$$\tilde{x} = \sum_{k=0}^{N} w_i x_i,$$

where the $\{w_i\}$'s are some weights, normalized such that $\sum_{k=0}^{N} w_i = 1$, define the averaging procedure. The simplest case corresponds to:

$$w_N = w_{N-1} = \cdots = w_{N-K+1} = \frac{1}{K}; \qquad w_k = 0 \quad (k < N - K + 1).$$

where the average is taken over the last K days of the option life. One however consider more complicated situations, for example an exponential ($(w_k \propto s^{N-k})$. The wealth balance then contains the modified pay-off: m $x_s, 0)$, or more generally $\mathcal{Y}(\tilde{x})$. The first problem therefore concerns the stati

$\tilde{x}$. As we shall see, this problem is very similar to the case encountered in Chapter 4 where the volatility is time dependent. Indeed, one has:

$$\sum_{k=0}^{N} w_i x_i = \sum_{k=0}^{N} w_i \left(\sum_{\ell=0}^{i-1} \delta x_\ell + x_0 \right) = x_0 + \sum_{k=0}^{N-1} \gamma_k \delta x_k, \tag{5.33}$$

where

$$\gamma_k \equiv \sum_{i=k+1}^{N} w_i. \tag{5.34}$$

Said differently, everything goes as if the price did not vary by an amount δx_k, but by an amount $\delta y_k = \gamma_k \delta x_k$, distributed as:

$$\frac{1}{\gamma_k} P_1 \left(\frac{\delta y_k}{\gamma_k} \right). \tag{5.35}$$

In the case of Gaussian fluctuations of variance $D\tau$, one thus finds:[2]

$$P(\tilde{x}, N|x_0, 0) = \frac{1}{\sqrt{2\pi \tilde{D} N \tau}} \exp\left[-\frac{(\tilde{x} - x_0)^2}{2\tilde{D}N\tau} \right], \tag{5.36}$$

where

$$\tilde{D} = \frac{D}{N} \sum_{k=0}^{N-1} \gamma_k^2. \tag{5.37}$$

More generally, $P(\tilde{x}, N|x_0, 0)$ is the Fourier transform of

$$\prod_{k=0}^{N-1} \hat{P}_1(\gamma_k z). \tag{5.38}$$

This information is sufficient to fix the option price (in the limit where the average return is very small) through:

$$\mathcal{C}_{\text{asi}}[x_0, x_s, N] = \int_{x_s}^{\infty} (\tilde{x} - x_s) P(\tilde{x}, N|x_0, 0) \, d\tilde{x}. \tag{5.39}$$

In order to fix the optimal strategy, one must however calculate the following quantity:

$$P(\tilde{x}, N|x, k) \langle \delta x_k \rangle |_{(x,k) \to (\tilde{x}, N)}, \tag{5.40}$$

conditioned to a certain terminal value for $\tilde{x}$ (cf. Eq. (4.74)). The general calculation is given in Appendix D. For a small kurtosis, the optimal strategy reads:

$$\phi_k^{N*}(x) = \frac{\partial \mathcal{C}_{\text{asi}}[x, x_s, N-k]}{\partial x} + \frac{\kappa_1 D\tau}{6} \gamma_k^2 \frac{\partial^3 \mathcal{C}_{\text{asi}}[x, x_s, N-k]}{\partial x^3}. \tag{5.41}$$

[2] The case of a multiplicative process is more involved: see, e.g. H. Gemam, M. Yor, Bessel processes, Asian Options and Perpetuities, *Mathematical Finance*, 3, 349 (1993).

Note that if the instant of time 'k' is outside the averaging period, one has $\gamma_k = 1$ (since $\sum_{i>k} w_i = 1$), and the formula, Eq. (4.80), is recovered. If on the contrary k gets closer to maturity, γ_k diminishes as does the correction term.

5.2.4 'American' options

We have up to now focused our attention on 'European'-type options, which can only be exercised on the day of expiry. In reality, most traded options on organized markets can be exercised at any time between the emission date and the expiry date: by definition, these are called 'American' options. It is obvious that the price of American options must greater or equal to the price of a European option with the same maturity and strike price, since the contract is *a priori* more favourable to the buyer. The pricing problem is therefore more difficult, since the writer of the option must first determine the optimal strategy that the buyer can follow in order to fix a fair price. Now, in the absence of dividends, the optimal strategy for the buyer of a call option is to keep it until the expiry date, thereby converting *de facto* the option into a European option. Intuitively, this is due to the fact that the average $\langle \max(x_N - x_s, 0) \rangle$ grows with N, hence the average pay-off is higher if one waits longer. The argument can be more convincing as follows. Let us define a 'two-shot' option, of strike x_s, which can only be exercised at times N_1 and $N_2 > N_1$ only.[3] At time N_1, the buyer of the option may choose to exercise a fraction $f(x_1)$ of the option, which in principle depends on the current price of the underlying x_1. The remaining part of the option can then be exercised at time N_2. What is the average profit $\langle \mathcal{G} \rangle$ of the buyer at time N_2?

Considering the two possible cases, one obtains:

$$\langle \mathcal{G} \rangle = \int_{x_s}^{+\infty} (x_2 - x_s) \, dx_2 \int P(x_2, N_2|x_1, N_1)[1 - f(x_1)] P(x_1, N_1|x_0, 0) \, dx_1$$
$$+ \int_{x_s}^{+\infty} f(x_1)(x_1 - x_s) e^{r\tau(N_2 - N_1)} P(x_1, N_1|x_0, 0) \, dx_1, \tag{5.42}$$

which can be rewritten as:

$$\langle \mathcal{G} \rangle = \mathcal{C}[x_0, x_s, N_2] e^{r\tau N_2} + \int_{x_s}^{\infty} f(x_1) \times P(x_1, N_1|x_0, 0) \, (x_1 - x_s - \mathcal{C}[x_1, x_s, N_2 - N_1]) \, e^{r\tau(N_2 - N_1)} \, dx_1. \tag{5.43}$$

The last expression means that if the buyer exercises a fraction $f(x_1)$ of his option, he pockets immediately the difference $x_1 - x_s$, but loses *de facto* his option, which is worth $\mathcal{C}[x_1, x_s, N_2 - N_1]$.

[3] Options that can be exercised at certain specific dates (more than one) are called 'Bermudan' options.

The optimal strategy, such that $\langle \mathcal{G} \rangle$ is maximum, therefore consists in choosing $f(x_1)$ equal to 0 or 1, according to the sign of $x_1 - x_s - \mathcal{C}[x_1, x_s, N_2 - N_1]$. Now, this difference is always negative, whatever x_1 and $N_2 - N_1$. This is due to the put–call parity relation (cf. Eq. (5.26)):

$$\mathcal{C}^\dagger[x_1, x_s, N_2 - N_1] = \mathcal{C}[x_1, x_s, N_2 - N_1] - (x_1 - x_s) - x_s(1 - e^{-r\tau(N_2 - N_1)}). \quad (5.44)$$

Since $\mathcal{C}^\dagger \geq 0$, $\mathcal{C}[x_1, x_s, N_2 - N_1] - (x_1 - x_s)$ is also greater or equal to zero.

The optimal value of $f(x_1)$ is thus zero; said differently the buyer should wait until maturity to exercise his option to maximize his average profit. This argument can be generalized to the case where the option can be exercised at any instant $N_1, N_2, \ldots, N_n$ with n arbitrary.

Note however that choosing a non-zero f increases the total probability of exercising the option, but reduces the average profit! More precisely, the total probability to reach x_s before maturity is twice the probability to exercise the option at expiry (if the distribution of δx is even, see Section 3.1.3). OTC American options are therefore favourable to the writer of the option, since some buyers might be tempted to exercise before expiry.

It is interesting to generalize the problem and consider the case where the two strike prices x_{s1} and x_{s2} are different at times N_1 and N_2, in particular in the case where $x_{s1} < x_{s2}$. The average profit, Eq. (5.43), is then equal to (for $r = 0$):

$$\begin{aligned}\langle \mathcal{G} \rangle &= \mathcal{C}[x_0, x_{s2}, N_2] + \int_{x_s}^{\infty} f(x_1) P(x_1, N_1|x_0, 0) \\ &\quad \times (x_1 - x_{s1} - \mathcal{C}[x_1, x_{s2}, N_2 - N_1])\, dx_1. \end{aligned} \quad (5.45)$$

The equation

$$x^* - x_{s1} - \mathcal{C}[x^*, x_{s2}, N_2 - N_1] = 0 \quad (5.46)$$

then has a non-trivial solution, leading to $f(x_1) = 1$ for $x_1 > x^*$. The average profit of the buyer therefore increases, in this case, upon an early exercise of the option.

American puts

Naively, the case of the American puts looks rather similar to that of the calls, and these should therefore also be equivalent to European puts. This is not the case for the following reason.[4] Using the same argument as above, one finds that the average profit associated to a 'two-shot' put option with exercise dates N_1, N_2 is given by:

$$\begin{aligned}\langle \mathcal{G}^\dagger \rangle &= \mathcal{C}^\dagger[x_0, x_s, N_2] e^{r\tau N_2} + \int_{-\infty}^{x_s} f(x_1) P(x_1, N_1|x_0, 0) \\ &\quad \times \left(x_s - x_1 - \mathcal{C}^\dagger[x_1, x_s, N_2 - N_1]\right) e^{r\tau(N_2 - N_1)}\, dx_1. \end{aligned} \quad (5.47)$$

[4] The case of American calls with non-zero dividends is similar to the case discussed here.

Now, the difference $(x_s - x_1 - \mathcal{C}^\dagger[x_1, x_s, N_2 - N_1])$ can be transformed, using the put–call parity, as:

$$x_s[1 - e^{-r\tau(N_2 - N_1)}] - \mathcal{C}[x_1, x_s, N_2 - N_1]. \quad (5.48)$$

This quantity may become positive if $\mathcal{C}[x_1, x_s, N_2 - N_1]$ is very small, which corresponds to $x_s \gg x_1$ (Puts deep in the money). The smaller the value of r, the larger should be the difference between x_1 and x_s, and the smaller the probability for this to happen. If $r = 0$, the problem of American puts is identical to that of the calls.

In the case where the quantity (5.48) becomes positive, an 'excess' average profit $\delta\mathcal{G}$ is generated, and represents the extra premium to be added to the price of the European put to account for the possibility of an early exercise. Let us finally note that the price of the American put $\mathcal{C}^\dagger_{am}$ is necessarily always larger or equal to $x_s - x$ (since this would be the immediate profit), and that the price of the 'two-shot' put is a lower bound to $\mathcal{C}^\dagger_{am}$.

The perturbative calculation of $\delta\mathcal{G}$ (and thus of the 'two-shot' option) in the limit of small interest rates is not very difficult. As a function of N_1, $\delta\mathcal{G}$ reaches a maximum between $N_2/2$ and N_2. For an at-the-money put such that $N_2 = 100$, $r = 5\%$ annual, $\sigma = 1\%$ per day and $x_0 = x_s = 100$, the maximum is reached for $N_1 \simeq 80$ and the corresponding $\delta\mathcal{G} \simeq 0.15$. This must be compared with the price of the European put, which is $\mathcal{C}^\dagger \simeq 4$. The possibility of an early exercise leads in this case to a 5% increase of the price of the option.

More generally, when the increments are independent and of average zero, one can obtain a numerical value for the price of an American put $\mathcal{C}^\dagger_{am}$ by iterating backwards the following exact equation:

$$\mathcal{C}^\dagger_{am}[x, x_s, N+1] = \max\left(x_s - x, e^{-r\tau} \int P_1(\delta x) \mathcal{C}^\dagger_{am}[x + \delta x, x_s, N]\, d\delta x\right). \quad (5.49)$$

This equation means that the put is worth the average value of tomorrow's price if it is not exercised today ($\mathcal{C}^\dagger_{am} > x_s - x$), or $x_s - x$ if it is immediately exercised. Using this procedure, we have calculated the price of a European, American and 'two-shot' option of maturity 100 days (Fig. 5.1). For the 'two-shot' option, the optimal value of N_1 as a function of the strike is shown in the inset.

5.2.5 *'Barrier' options*

Let us now turn to another family of options, called 'barrier' options, which are such that if the price of the underlying x_k reaches a certain 'barrier' value x_b during the lifetime of the option, the option is lost. (Conversely, there are options that are only activated if the value x_b is reached.) This clause leads to cheaper options, which can be more attractive to the investor. Also, if $x_b > x_s$, the writer of the option limits his possible losses to $x_b - x_s$. What is the probability $P_b(x, N|x_0, 0)$ for the final value of the underlying to be at x, conditioned to the fact that the price has not reached the barrier value x_b?

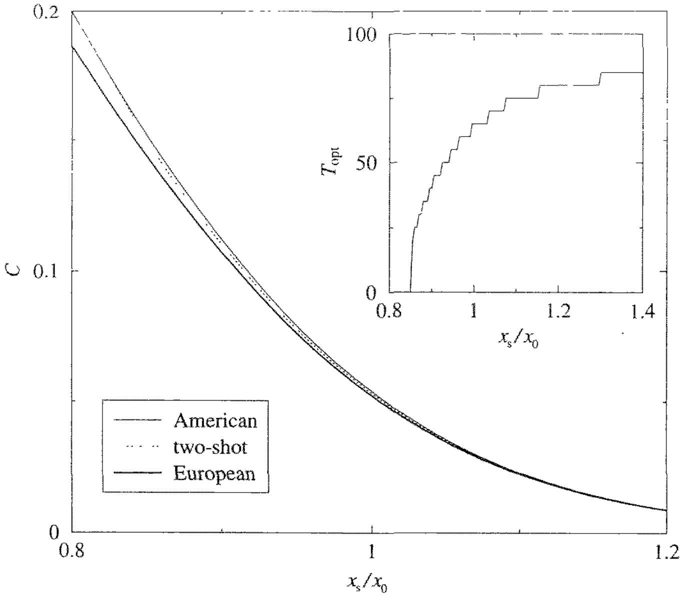

Fig. 5.1. Price of a European, American and 'two-shot' put option as a function of the strike, for a 100-days maturity and a daily volatility of 1% and $r = 1\%$. The top curve is the American price, while the bottom curve is the European price. In the inset is shown the optimal exercise time N_1 as a function of the strike for the 'two-shot' option.

In some cases, it is possible to give an exact answer to this question, using the so-called method of images. Let us suppose that for each time step, the price x can only change by an discrete amount, ± 1 tick. The method of images is explained graphically in Figure 5.2: one can notice that all the trajectories going through x_b between $k = 0$ and $k = N$ has a 'mirror' trajectory, with a statistical weight precisely equal (for $m = 0$) to the one of the trajectory one wishes to exclude. It is clear that the conditional probability we are looking for is obtained by subtracting the weight of these image trajectories:

$$P_b(x, N|x_0, 0) = P(x, N|x_0, 0) - P(x, N|2x_b - x_0, 0). \tag{5.50}$$

In the general case where the variations of x are not limited to $0, \pm 1$, the previous argument fails, as one can easily be convinced by considering the case where δx takes the values ± 1 and ± 2. However, if the possible variations of the price during the time τ are small, the error coming from the uncertainty about the exact crossing time is small, and leads to an error on the price $\mathcal{C}_b$ of the barrier option on the order of $\langle|\delta x|\rangle$ times the total probability of ever touching the barrier. Discarding this correction, the price of barrier options reads:

$$\begin{aligned}\mathcal{C}_b[x_0, x_s, N] &= \int_{x_s}^{\infty} (x - x_s)[P(x, N|x_0, 0) - P(x, N|2x_b - x_0, 0)]\,\mathrm{d}x \\ &\equiv \mathcal{C}[x_0, x_s, N] - \mathcal{C}[2x_b - x_0, x_s, N] \end{aligned} \tag{5.51}$$

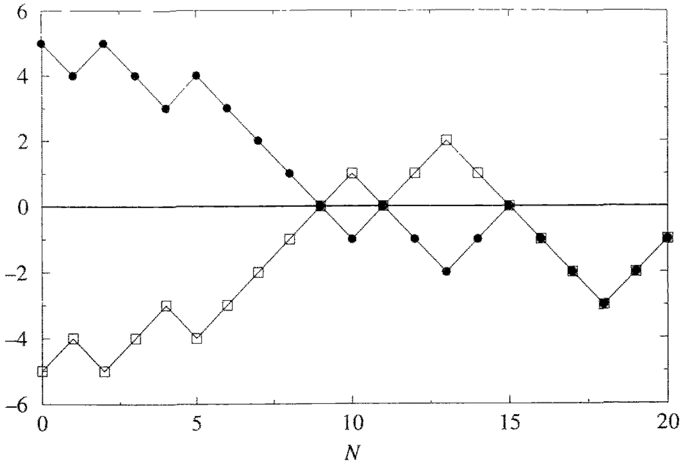

Fig. 5.2. Illustration of the method of images. A trajectory, starting from the point $x_0 = -5$ and reaching the point $x_{20} = -1$ can either touch or avoid the 'barrier' located at $x_b = 0$. For each trajectory touching the barrier, as the one shown in the figure (squares), there exists one single trajectory (circles) starting from the point $x_0 = +5$ and reaching the same final point – only the last section of the trajectory (after the last crossing point) is common to both trajectories. In the absence of bias, these two trajectories have exactly the same statistical weight. The probability of reaching the final point without crossing $x_b = 0$ can thus be obtained by subtracting the weight of the image trajectories. Note that the whole argument is wrong if jump sizes are not constant (for example when $\delta x = \pm 1$ or ± 2).

($x_b < x_0$), or

$$\mathcal{C}_b[x_0, x_s, N] = \int_{x_s}^{x_b} (x - x_s)\,[P(x, N|x_0, 0) - P(x, N|2x_b - x_0, 0)]\,\mathrm{d}x, \tag{5.52}$$

($x_b > x_s$); the option is worthless whenever $x_0 < x_b < x_s$.

One can also find 'double barrier' options, such that the price is constrained to remain within a certain channel $x_b^- < x < x_b^+$, or else the option vanishes. One can generalize the method of images to this case. The images are now successive reflections of the starting point x_0 in the two parallel 'mirrors' x_b^-, x_b^+.

Other types of option

One can find many other types of option, which we shall not discuss further. Some options, for example, are calculated on the maximum value of the price of the underlying reached during a certain period. It is clear that in this case, a Gaussian or log-normal model is particularly inadequate, since the price of the option is governed by extreme events. Only an adequate treatment of the tails of the distribution can allow us to price this type of option correctly.

5.3 The 'Greeks' and risk control

The 'Greeks', which is the traditional name given by professionals to the derivative of the price of an option with respect to the price of the underlying, the volatility, etc., are often used for local risk control purposes. Indeed, if one assumes that the underlying asset does not vary too much between two instants of time t and $t+\tau$, one may expand the variation of the option price in Taylor series:

$$\delta C = \Delta \delta x + \frac{1}{2}\Gamma(\delta x)^2 + \mathcal{V}\delta\sigma + \Theta\tau, \tag{5.53}$$

where δx is the change of price of the underlying. If the option is hedged by simultaneously selling a proportion ϕ of the underlying asset, one finds that the change of the portfolio value is, to this order:

$$\delta W = (\Delta - \phi)\delta x + \frac{1}{2}\Gamma(\delta x)^2 + \mathcal{V}\delta\sigma + \Theta\tau. \tag{5.54}$$

Note that the Black–Scholes (or rather, Bachelier) equation is recovered by setting $\phi^* = \Delta$, $\delta\sigma \equiv 0$, and by recalling that for a continuous-time Gaussian process, $(\delta x)^2 \equiv D\tau$ (see Section 4.5.2). In this case, the portfolio does not change with time ($\delta W = 0$), provided that $\Theta = -D\Gamma/2$, which is precisely Eq. (4.51) in the limit $\tau \to 0$.

In reality, due to the non-Gaussian nature of δx, the large risk corresponds to cases where $\Gamma(\delta x)^2 \gg |\Theta\tau|$. Assuming that one chooses to follow the Δ-hedge procedure (which is in general suboptimal, see Section 4.4.3 above), one finds that the fluctuations of the price of the underlying leads to an *increase* in the value of the portfolio of the buyer of the option (since $\Gamma > 0$). Losses can only occur if the implied volatility of the underlying decreases. If δx and $\delta\sigma$ are uncorrelated (which is in general not true), one finds that the 'instantaneous' variance of the portfolio is given by:

$$\langle(\delta W)^2\rangle = \frac{3+\kappa_1}{4}\Gamma^2\langle\delta x^2\rangle^2 + \mathcal{V}^2\langle\delta\sigma^2\rangle, \tag{5.55}$$

where κ_1 is the kurtosis of δx. For an at-the-money option of maturity T, one has:

$$\Gamma \sim \frac{1}{\sigma x_0\sqrt{T}} \qquad \mathcal{V} \sim x_0\sqrt{T}. \tag{5.56}$$

Typical values are, on the scale of τ = one day, $\kappa_1 = 3$ and $\delta\sigma \sim \sigma$. The Γ contribution to risk is therefore on the order of $\sigma x_0 \tau/\sqrt{T}$. This is equal to the typical fluctuations of the underlying contract multiplied by $\sqrt{\tau/T}$, or else the price of the option reduced by a factor $N = T/\tau$. The Vega contribution is much larger for long maturities, since it is of order of the price of the option itself.

5.4 Value-at-risk for general non-linear portfolios (*)

A very important issue for the control of risk of complex portfolios, which involves many non-linear assets, is to be able to estimate its value-at-risk reliably. This is a difficult problem, since both the non-Gaussian nature of the fluctuations of the underlying assets and the non-linearities of the price of the derivatives must be dealt with. A solution, which is very costly in terms of computation time and not very precise, is the use of Monte-Carlo simulations. We shall show in this section that in the case where the fluctuations of the 'explicative variables' are strong (a more precise statement will be made below), an approximate formula can be obtained for the value-at-risk of a general non-linear portfolio.

Let us assume that the variations of the value of the portfolio can be written as a function $\delta f(e_1, e_2, \ldots, e_M)$ of a set of M independent random variables e_a, $a = 1, \ldots, M$, such that $\langle e_a\rangle = 0$ and $\langle e_a e_b\rangle = \delta_{a,b}\sigma_a^2$. The sensitivity of the portfolio to these 'explicative variables' can be measured as the derivatives of the value of the portfolio with respect to the e_a. We shall therefore introduce the Δ's and Γ's as:

$$\Delta_a = \frac{\partial f}{\partial e_a} \qquad \Gamma_{a,b} = \frac{\partial^2 f}{\partial e_a \partial e_b}. \tag{5.57}$$

We are interested in the probability for a large fluctuation δf^* of the portfolio. We will surmise that this is due to a particularly large fluctuation of one explicative factor, say $a = 1$, that we will call the dominant factor. This is not always true, and depends on the statistics of the fluctuations of the e_a. A condition for this assumption to be true will be discussed below, and requires in particular that the tail of the dominant factor should not decrease faster than an exponential. Fortunately, this is a good assumption in financial markets.

The aim is to compute the value-at-risk of a certain portfolio, i.e. the value δf^* such that the probability that the variation of f exceeds δf^* is equal to a certain probability p: $\mathcal{P}_>(\delta f^*) = p$. Our assumption about the existence of a dominant factor means that these events correspond to a market configuration where the fluctuation δe_1 is large, whereas all other factors are relatively small. Therefore, the large variations of the portfolio can be approximated as:

$$\delta f(e_1, e_2, \ldots, e_M) = \delta f(e_1) + \sum_{a=2}^{M}\Delta_a e_a + \frac{1}{2}\sum_{a,b=2}^{M}\Gamma_{a,b}e_a e_b, \tag{5.58}$$

where $\delta f(e_1)$ is a shorthand notation for $\delta f(e_1, 0, \ldots, 0)$. Now, we use the fact that:

$$\mathcal{P}_>(\delta f^*) = \int P(e_1, e_2, \ldots, e_M)\Theta\left[\delta f(e_1, e_2, \ldots, e_M) - \delta f^*\right]\prod_{a=1}^{M} de_a. \tag{5.59}$$

where $\Theta(x > 0) = 1$ and $\Theta(x < 0) = 0$. Expanding the Θ function to second order leads to:

$$\Theta(\delta f(e_1) - \delta f^*) + \left[\sum_{a=2}^{M} \Delta_a e_a + \frac{1}{2}\sum_{a=2}^{M} \Gamma_{a,b} e_a e_b\right] \delta(\delta f(e_1) - \delta f^*)$$
$$+ \frac{1}{2}\sum_{a,b=2}^{M} \Delta_a \Delta_b e_a e_b \delta'(\delta f(e_1) - \delta f^*), \tag{5.60}$$

where δ' is the derivative of the δ-function with respect to δf. In order to proceed with the integration over the variables e_a in Eq. (5.59), one should furthermore note the following identity:

$$\delta(\delta f(e_1) - \delta f^*) = \frac{1}{\Delta_1^*}\delta(e_1 - e_1^*), \tag{5.61}$$

where e_1^* is such that $\delta f(e_1^*) = \delta f^*$, and Δ_1^* is computed for $e_1 = e_1^*$, $e_{a>1} = 0$. Inserting the above expansion of the Θ function into Eq. (5.59) and performing the integration over the e_a then leads to:

$$\mathcal{P}_>(\delta f^*) = \mathcal{P}_>(e_1^*) + \sum_{a=2}^{M} \frac{\Gamma^*_{a,a}\sigma_a^2}{2\Delta_1^*} P(e_1^*) - \sum_{a=2}^{M} \frac{\Delta_a^{*2}\sigma_a^2}{2\Delta_1^{*2}} \left(P'(e_1^*) + \frac{\Gamma^*_{1,1}}{\Delta_1^*} P(e_1^*)\right), \tag{5.62}$$

where $P(e_1)$ is the probability distribution of the first factor, defined as:

$$P(e_1) = \int P(e_1, e_2, \ldots, e_M) \prod_{a=2}^{M} \mathrm{d}e_a. \tag{5.63}$$

In order to find the value-at-risk δf^*, one should thus solve Eq. (5.62) for e_1^* with $\mathcal{P}_>(\delta f^*) = p$, and then compute $\delta f(e_1^*, 0, \ldots, 0)$. Note that the equation is not trivial since the Greeks must be estimated at the solution point e_1^*.

Let us discuss the general result, Eq. (5.62), in the simple case of a linear portfolio of assets, such that no convexity is present: the Δ_a's are constant and the $\Gamma_{a,a}$'s are all zero. The equation then takes the following simpler form:

$$\mathcal{P}_>(e_1^*) - \sum_{a=2}^{M} \frac{\Delta_a^2 \sigma_a^2}{2\Delta_1^2} P'(e_1^*) = p. \tag{5.64}$$

Naively, one could have thought that in the dominant factor approximation, the value of e_1^* would be the value-at-risk value of e_1 for the probability p, defined as:

$$\mathcal{P}_>(e_{1,\mathrm{VaR}}) = p. \tag{5.65}$$

However, the above equation shows that there is a correction term proportional to $P'(e_1^*)$. Since the latter quantity is negative, one sees that e_1^* is actually larger than

$e_{1,\mathrm{VaR}}$, and therefore $\delta f^* > \delta f(e_{1,\mathrm{VaR}})$. This reflects the effect of all other factors, which tend to increase the value-at-risk of the portfolio.

The result obtained above relies on a second-order expansion; when are higher-order corrections negligible? It is easy to see that higher-order terms involve higher-order derivatives of $P(e_1)$. A condition for these terms to be negligible in the limit $p \to 0$, or $e_1^* \to \infty$, is that the successive derivatives of $P(e_1)$ become smaller and smaller. This is true provided that $P(e_1)$ decays more slowly than exponentially, for example as a power-law. On the contrary, when $P(e_1)$ decays faster than exponentially (for example in the Gaussian case), then the expansion proposed above completely loses its meaning, since higher and higher corrections become dominant when $p \to 0$. This is expected: in a Gaussian world, a large event results from the accidental superposition of many small events, whereas in a power-law world, large events are associated to one single large fluctuation which dominates over all the others. The case where $P(e_1)$ decays as an exponential is interesting, since it is often a good approximation for the tail of the fluctuations of financial assets. Taking $P(e_1) \simeq \alpha_1 \exp -\alpha_1 e_1$, one finds that e_1^* is the solution of:

$$\mathrm{e}^{-\alpha_1 e_1^*} \left[1 - \sum_{a=2}^{M} \frac{\Delta_a^2 \alpha_1^2 \sigma_a^2}{2\Delta_1^2}\right] = p. \tag{5.66}$$

Since one has $\sigma_1^2 \propto \alpha_1^{-2}$, the correction term is small provided that the variance of the portfolio generated by the dominant factor is much larger than the sum of the variance of all other factors.

Coming back to Eq. (5.62), one expects that if the dominant factor is correctly identified, and if the distribution is such that the above expansion makes sense, an approximate solution is given by $e_1^* = e_{1,\mathrm{VaR}} + \epsilon$, with:

$$\epsilon \simeq \sum_{a=2}^{M} \frac{\Gamma_{a,a}\sigma_a^2}{2\Delta_1} - \sum_{a=2}^{M} \frac{\Delta_a^2 \sigma_a^2}{2\Delta_1^2} \left(\frac{P'(e_{1,\mathrm{VaR}})}{P(e_{1,\mathrm{VaR}})} + \frac{\Gamma_{1,1}}{\Delta_1}\right), \tag{5.67}$$

where now all the Greeks at estimated at $e_{1,\mathrm{VaR}}$.

In some cases, it appears that a 'one-factor' approximation is not enough to reproduce the correct VaR value. This can be traced back to the fact that there are actually other different dangerous market configurations which contribute to the VaR. The above formalism can however easily be adapted to the case where two (or more) dangerous configurations need to be considered. The general equations read:

$$\mathcal{P}_{>a} = \mathcal{P}_>(e_a^*) + \sum_{b\neq a}^{M} \frac{\Gamma^*_{b,b}\sigma_b^2}{2\Delta_a^*} P(e_a^*) - \sum_{b\neq a}^{M} \frac{\Delta_2^{*2}\sigma_b^*}{2\Delta_a^{*2}} \left(P'(e_a^*) + \frac{\Gamma^*_{a,a}}{\Delta_a^*} P(e_a^*)\right), \tag{5.68}$$

where $a = 1, \ldots, K$ are the K different dangerous factors. The e_a^* and therefore

δf^*, are determined by the following K conditions:

$$\delta f^*(e_1^*) = \delta f^*(e_2^*) = \cdots = \delta f^*(e_K^*) \qquad \mathcal{P}_{>1} + \mathcal{P}_{>2} + \cdots + \mathcal{P}_{>K} = p. \quad (5.69)$$

5.5 Risk diversification (*)

We have put the emphasis on the fact that for real world options, the Black–Scholes divine surprise – i.e. the fact that the risk is zero – does not occur, and a non-zero residual risk remains. One can ask whether this residual risk can be reduced further by including other assets in the hedging portfolio. Buying stocks other than the underlying to hedge an option can be called an 'exogenous' hedge. A related question concerns the hedging of a 'basket' option, the pay-off of which being calculated on a linear superposition of different assets. A rather common example is that of 'spread' options, which depend on the *difference* of the price between two assets (for example the difference between the Nikkei and the S&P 500, or between the British and German interest rates, etc.). An interesting conclusion is that in the Gaussian case, an exogenous hedge increases the risk. An exogenous hedge is only useful in the presence of non-Gaussian effects. Another possibility is to hedge some options using different options; in other words, one can ask how to optimize a whole 'book' of options such that the global risk is minimum.

'Portfolio' options and 'exogenous' hedging

Let us suppose that one can buy M assets X^i, $i = 1, \ldots, M$, the price of which being x_k^i at time k. As in Chapter 3, we shall suppose that these assets can be decomposed over a basis of independent factors E^a:

$$x_k^i = \sum_{a=1}^{M} O_{ia} e_k^a. \quad (5.70)$$

The E^a are independent, of unit variance, and of distribution function P_a. The correlation matrix of the fluctuations, $\langle \delta x^i \delta x^j \rangle$ is equal to $\sum_a O_{ia} O_{ja} = [\mathbf{OO}^\dagger]_{ij}$.

One considers a general option constructed on a linear combination of all assets, such that the pay-off depends on the value of

$$\tilde{x} = \sum_i f_i x^i \quad (5.71)$$

and is equal to $\mathcal{Y}(\tilde{x}) = \max(\tilde{x} - x_s, 0)$. The usual case of an option on the asset X^1 thus corresponds to $f_i = \delta_{i,1}$. A spread option on the difference $X^1 - X^2$ corresponds to $f_i = \delta_{i,1} - \delta_{i,2}$, etc. The hedging portfolio at time k is made of all the different assets X^i, with weight ϕ_k^i. The question is to determine the optimal composition of the portfolio, ϕ_k^{i*}.

Following the general method explained in Section 4.3.3, one finds that the part of the risk which depends on the strategy contains both a linear and a quadratic term in the ϕ's. Using the fact that the E^a are independent random variables, one can compute the functional derivative of the risk with respect to all the $\phi_k^i(\{x\})$. Setting this functional derivative to zero leads to:[5]

$$\sum_j [\mathbf{OO}^\dagger]_{ij} \phi_k^j = \int \hat{\mathcal{Y}}(z) \left[\prod_b \hat{P}_b(z \sum_j f_j O_{jb}) \right] \times \sum_a \frac{O_{ia}}{\sum_j f_j O_{ja}} \frac{\partial}{\partial \mathrm{i}z} \log \hat{P}_a \left(z \sum_j f_j O_{ja} \right) \mathrm{d}z. \quad (5.72)$$

Using the cumulant expansion of P_a (assumed to be even), one finds that:

$$\frac{\partial}{\partial \mathrm{i}z} \log \hat{P}_a \left(z \sum_j f_j O_{ja} \right) = \mathrm{i}z \left(\sum_j f_j O_{ja} \right)^2 - \mathrm{i}\frac{z^3}{6} \kappa_a \left(\sum_j f_j O_{ja} \right)^4 + \cdots. \quad (5.73)$$

The first term combines with

$$\sum_a \frac{O_{ia}}{\sum_j f_j O_{ja}}, \quad (5.74)$$

to yield:

$$\mathrm{i}z \sum_{aj} O_{ia} O_{ja} f_j \equiv \mathrm{i}z [\mathbf{OO}^\dagger \cdot f]_a, \quad (5.75)$$

which finally leads to the following simple result:

$$\phi_k^{i*} = f_i \mathcal{P}[\{x_k^i\}, x_s, N - k] \quad (5.76)$$

where $\mathcal{P}[\{x_k^i\}, x_s, N - k]$ is the probability for the option to be exercised, calculated at time k. In other words, in the Gaussian case ($\kappa_a \equiv 0$) the optimal portfolio is such that the proportion of asset i precisely reflects the weight of i in the basket on which the option is constructed. In particular, in the case of an option on a single asset, the hedging strategy is not improved if one includes other assets, even if these assets are correlated with the former.

However, this conclusion is only correct in the case of Gaussian fluctuations and does not hold if the kurtosis is non-zero.[6] *In this case, an extra term appears, given by:*

$$\delta \phi_k^{i*} = \frac{1}{6} \left[\sum_{ja} \kappa_a [\mathbf{OO}^\dagger]_{ij}^{-1} O_{ja} \sum_l f_l O_{la}^3 \right] \frac{\partial P(\tilde{x}, x_s, N - k)}{\partial x_s}. \quad (5.77)$$

This correction is not, in general, proportional to f_i, and therefore suggests that, in some cases, an exogenous hedge can be useful. However, one should note that this correction is small for at-the-money options ($\tilde{x} = x_s$), since $\partial P(\tilde{x}, x_s, N - k)/\partial x_s = 0$.

[5] In the following, i denotes the unit imaginary number, except when it appears as a subscript, in which case it is an asset label.

[6] The case of Lévy fluctuations is also such that an exogenous hedge is useless.

Option portfolio

Since the risk associated with a single option is in general non-zero, the global risk of a portfolio of options ('book') is also non-zero. Suppose that the book contains p_i calls of 'type' i (i therefore contains the information of the strike x_{si} and maturity T_i). The first problem to solve is that of the hedging strategy. In the absence of volatility risk, it is not difficult to show that the optimal hedge for the book is the linear superposition of the optimal strategies for each individual option:

$$\phi^*(x,t) = \sum_i p_i \phi_i^*(x,t). \tag{5.78}$$

The residual risk is then given by:

$$\mathcal{R}^{*2} = \sum_{i,j} p_i p_j C_{ij}, \tag{5.79}$$

where the 'correlation matrix' C is equal to:

$$\begin{aligned} C_{ij} &= \big\langle \max(x(T_i) - x_{si}, 0)\max(x(T_j) - x_{sj}, 0)\big\rangle \\ &\quad - \mathcal{C}_i\mathcal{C}_j - D\tau \sum_{k=0}^{N-1} \langle \phi_i^*(x, k\tau)\phi_j^*(x, k\tau)\rangle \end{aligned} \tag{5.80}$$

where $\mathcal{C}_i$ is the price of the option i. If the constraint on the p_i's is of the form $\sum_i p_i = 1$, the optimum portfolio is given by:

$$p_i^* = \frac{\sum_j C_{ij}^{-1}}{\sum_{i,j} C_{ij}^{-1}} \tag{5.81}$$

(remember that by assumption the mean return associated to an option is zero).

Let us finally note that we have not considered, in the above calculation, the risk associated with volatility fluctuations, which is rather important in practice. It is a common practice to try to hedge this volatility risk using other types of options (for example, an exotic option can be hedged using a 'plain vanilla' option). A generalization of the Black–Scholes argument (assuming that option prices themselves follow a Gaussian process, which is far from being the case) suggests that the optimal strategy is to hold a fraction

$$\tilde{\Delta} = \frac{\partial \mathcal{C}_2}{\partial \sigma} \Big/ \frac{\partial \mathcal{C}_1}{\partial \sigma} \tag{5.82}$$

of options of type 2 to hedge the volatility risk associated with an option of type 1. Using the formalism established in Chapter 4, one could work out the correct hedging strategy, taking into account the non-Gaussian nature of the price variations of options.

5.6 References

More on options, exotic options

J. C. Hull, *Futures, Options and Other Derivatives*, Prentice Hall, Upper Saddle River, NJ, 1997.
P. W. Wilmott, J. N. Dewynne, S. D. Howison, *Option Pricing: Mathematical Models and Computation*, Cambridge University Press, Cambridge, 1997.
N. Taleb, *Dynamical Hedging: Managing Vanilla and Exotic Options*, Wiley, New York, 1998.
I. Nelken (ed.), *The Handbook of Exotic Options*, Irwin, Chicago, IL, 1996.

Stochastic volatility models and volatility hedging

J. C. Hull, A. White, The pricing of options on assets with stochastic volatility, *Journal of Finance*, **XLII**, 281 (1987); *idem*, An analysis of the bias in option pricing caused by stochastic volatility, *Advances in Futures and Option Research*, **3**, 29 (1988).

Short glossary of financial terms

Arbitrage A trading strategy that generates profit without risk, from a zero initial investment.

Basis point Elementary price increment, equal to 10^{-4} in relative value.

Bid–ask spread Difference between the ask price (at which one can buy an asset) and the bid price (at which one can sell the same asset).

Bond (Zero coupon): Financial contract which pays a fixed value at a given date in the future.

Delta Derivative of the price of an option with respect to the current price of the underlying contract. This is equal to the optimal hedging strategy in the Black–Scholes world.

Drawdown Period of time during which the price of an asset is below its last historical peak.

Forward Financial contract under which the owner agrees to buy for a fixed price some asset, at a fixed date in the future.

Futures Same as a forward contract, but on an organized market. In this case, the contract is marked-to-market, and the owner pays (or receives) the marginal price change on a daily basis.

Gamma Second derivative of the price of an option with respect to the current price of the underlying contract. This is equal to the derivative of the optimal hedging strategy in the Black–Scholes world.

Hedging strategy A trading strategy allowing one to reduce, or sometimes to eliminate completely, the risk of a position.

Moneyness Describes the difference between the spot price and the strike price of an option. For a call, if this difference is positive [resp. negative], the option is said to be in-the-money [resp. out-of-the-money]. If the difference is zero, the option is at-the-money.

Option Financial contract allowing the owner to buy [or sell] at a fixed maximum [minimum] price (the strike price) some underlying asset in the future.

This contract protects its owner against a possible rise or fall in price of the underlying asset.

Over-the-counter This is said of a financial contract traded off market, say between two financial companies or banks. The price is then usually not publicly disclosed, at variance with organized markets.

Spot price The current price of an asset for immediate delivery, in contrast with, for example, its forward price.

Spot rate The value of the short-term interest rate.

Spread Difference in price between two assets, or between two different prices of the same asset – for example, the bid–ask spread.

Strike price Price at which an option can be exercised, see Option.

Vega Derivative of the price of an option with respect to the volatility of the underlying contract.

Value at Risk (VaR) Measure of the potential losses of a given portfolio, associated to a certain confidence level. For example, a 95% VaR corresponds to the loss level that has a 5% probability to be exceeded.

Volatility Standard deviation of an asset's relative price changes.

Index of symbols

Index